



HAL
open science

A glance at non-standard model and logics of uncertainty and vagueness

Didier Dubois, Henri Prade

► To cite this version:

Didier Dubois, Henri Prade. A glance at non-standard model and logics of uncertainty and vagueness. Jacques-Paul Dubucs. Philosophy of Probability, 56, Kluwer; Springer Netherlands, pp.169–222, 1993, Philosophical Studies Series book series (PSSP), 978-90-481-4301-6. <10.1007/978-94-015-8208-7_9>. <hal-04064854>

HAL Id: hal-04064854

<https://hal.science/hal-04064854v1>

Submitted on 13 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

A glance at non-standard models and logics of uncertainty and vagueness

Didier DUBOIS – Henri PRADE

Institut de Recherche en Informatique de Toulouse – I.R.I.T.

Université Paul Sabatier, 118 route de Narbonne

31062 TOULOUSE Cedex – FRANCE

Introduction

Historically, it is well known that the notion of probability emerged in the 17th century as a dual concept : chance, related to gaming problems, and subjective uncertainty, related to the question of reliability of testimonies. In the works of pioneers of probability theory, such as J. Bernoulli, chance, very soon connected to frequency of occurrence, was an additive notion but subjective probability was not so. However with the development of physical sciences, the non-additive side of probability was forgotten (see Shafer, 1978). So much so as 20th century researchers in decision theory have devoted much effort in the non-frequentist justification of additive probability as a model for subjective uncertainty in rational decision strategies.

The advent of the computer age has recently produced numerous works whose focus is the formal representation of various kinds of imperfect knowledge, with no particular emphasis on decision. These works seem to prompt a revival of non-additive probability. They also bridge the gap between uncertainty and very old notions of modality (such as possibility and necessity) already used by Aristotle and Middle-Ages scholars, notions which have been reintroduced with full strength in this century by modal logics.

This paper discusses recent approaches to the representation of imperfect information, as they have appeared in the last twenty years. The stress is on ideas rather than mathematics. Especially, it is pointed out that beside old questions such as the numerical representation of partial belief, new aspects of information imperfection are the topic of active research, such as vagueness and granularity. Moreover, while the problem of devising procedures for decision-

making in the face of imperfect information is still an important area, much research in Artificial Intelligence focuses on reasoning and methods for automating reasoning processes on the computer. The latter aims at devising intelligent information systems, as opposed to decision-support systems. Clearly issues pertaining to vagueness and granularity are present in information system research as much as issues related to uncertainty and incompleteness of available knowledge. Moreover, this trend has put forward a strong emphasis on logic, in non-standard forms, a concern that was not so common in past schools of uncertainty modelling, such as in statistics or decision theory.

The thrust of the paper is as follows : in the first part, traditional views of uncertainty based on probability are briefly recalled and taken as a starting point. Then several trends are reviewed that depart from these traditional views : they reflect concerns about other aspects of imperfect information such as imprecision, belief viewed as an ordering relation, lexical vagueness, and granularity. An analysis of how these aspects may interfere is provided ; it is also pointed out that these notions can often be considered from two points of view which can be called objective and subjective, respectively. In the second part the question of finding counterparts of well-established notions such as conditioning, Bayesian updating, information measures in the field of non-standard uncertainty models is evoked. In the third part, we try to classify various kind of logics induced by the various aspects of information imperfection. The distinction between degree of belief and what can be called degree of truth is recalled, and serves as a tool for distinguishing between logics of uncertainty and logics of vagueness. It is stressed that several types of logics of uncertainty can be envisaged.

I - Various Facets of Imperfect Information

I.1 - Probability and Knowledge Representation

The canonical setting of probability theory starts with a set of exhaustive and mutually exclusive outcomes, and considers a point-valued numerical additive set-function that assigns to each possible event its probability. As it is well known, there are several interpretations of probability (Fine, 1973). The three most usual interpretations are respectively a ratio of favorable outcomes over possible outcomes (with the game of chance paradigm), a frequency in the long run, and a degree of belief (usually measured by means of odds in a gambling experiment, following De Finetti (1937)). As Shafer (1990) pointed out these views are completely entwined and form the ideal picture of probability : indeed we claim that the frequency of heads in a coin-flipping experiment is 0.5, when the coin is fair; this state of facts dictates what fair odds are in a gambling game, and belief is then naturally related to odds. However the ideal picture of

probability assumes that much information is available, and tightly links belief and decision, through the setting of repeated experiments.

The problem of knowledge representation is quite different and relates only loosely to the ideal picture of probability. The problem of rational behavior in front of a random experiment is turned into the problem of faithfully representing evidence and drawing conclusions of interest from such representations. Evidence does not necessarily appear under the form of results of coin-flipping experiments : sometimes it takes the form of ill-observed statistics (e.g. the results of opinion polls), sometimes it consists of incomplete knowledge of frequencies. But it can appear under much poorer forms as an ordering over a set of events (some being more believed than others, because, for instance, the former correspond to the normal course of things while the latter does not), or a set of statements in natural language. In each case, there is some aspect of imperfection in information that is not met in the ideal picture of probability, namely incompleteness or imprecision (that naturally leads to interval-valued representations of uncertainty), lack of additive structure (e.g. counting outcomes make no sense), vagueness (that forbids to interpret events as subsets of outcomes in the usual sense), and granularity (induced by the fact that, for instance some outcomes are indiscernible from one another, in a way that depends upon the observer).

I.2 - Incomplete Knowledge

In its most elementary form, incomplete knowledge appears when some variable of interest is such that its true value is only known to belong to a subset of possible values. For instance John is known to be between 20 and 30 years old, the die throw produced an even number, etc.... More generally, it corresponds to a lack of completeness in the logical sense, i.e. information that does not allow to decide whether any proposition of interest is true or false. Let E be a subset of the set Ω of outcomes, such that all we know is that E has occurred. What can be said about the occurrence of other events A ? the uncertainty of they occurrence can be assessed in terms of possibility and necessity as follows :

- A necessarily occurred as soon as E implies A ($E \subseteq A$). The corresponding set-function is denoted N , and is such that $N(A) = 1$ if A necessarily occurs and 0 otherwise ;
- A possibly occurred as soon as E is consistent with A , i.e. $A \cap E \neq \emptyset$. The corresponding set-function is denoted Π and is such that $\Pi(A) = 1$ if A is consistent with E and 0 otherwise.

Π and N are respectively called possibility and necessity measures (Zadeh, 1978a ; Dubois and Prade, 1980, 1985c). Note that $N(A) = 0$ denotes a complete lack of certainty while $\Pi(A) = 0$ denotes impossibility. Possibility and necessity measures satisfy the remarkable duality

relationship

$$\prod(A) = 1 - N(\bar{A}) \quad (1)$$

where the overbar denotes complementation. This duality is also found in modal logics (Hughes and Cresswell, 1968) and goes back to Aristotles. It means that A is impossible if and only if its complement is necessary (i.e. necessarily true). \prod and N appear as natural ways of representing uncertainty in the presence of incomplete information. Some may find an average evaluation such as $|A \cap E| / |E|$ is a more flexible way of expressing the uncertainty of A (where $|\cdot|$ denotes cardinality). This quantity corresponds to a particular case of Carnap (1950)'s confirmation function, that evaluates to what extent E confirms A . This particular choice boils down to constructing a uniform probability density over E ; it presupposes that every realization of E has equal weight. That is exactly what possibility and necessity measures search to avoid.

Combining incomplete information with the canonical situation of probability theory leads to various kinds of extensions of probability measures that are also many-valued extensions of elementary possibility and necessity measures described above.

A first extension is to admit that only some kinds of events can be (or have been) observed, because observations are tainted with imprecision. Then the set of observable outcomes is a subset of the power set 2^Ω , say a family $\{E_1, \dots, E_n\}$ of non-empty sets, and probabilities $m(E_1), m(E_2), \dots, m(E_n)$ that sum to 1 are available. $m(E_i)$ can be the long-run frequency of *observing exactly* E_i , or the subjective probability of this observation (as derived from a betting behavior). Then letting \prod_i and N_i be possibility and necessity measures derived from E_i , uncertainty is quantified by means of expected possibility and necessity functions

$$\prod(A) = \sum_{i=1,n} m(E_i) \cdot \prod_i(A) = \sum_{E_i \cap A \neq \emptyset} m(E_i) \quad (2)$$

$$N(A) = \sum_{i=1,n} m(E_i) \cdot N_i(A) = \sum_{E_i \subseteq A} m(E_i) = 1 - \prod(\bar{A}) \quad (3)$$

\prod and N as defined above are respectively called plausibility and belief functions by Shafer (1976). Note that this type of model is not always related to imprecision of the observation device. It may correspond to a case when a classical probability paradigm can be defined over a set Ω of mutually exclusive and exhaustive answers to a question, but the question of interest (the answers of which we want to assess the uncertainty) is a related question, and this relationship is not one to one (Shafer,1990). Formally, this model is closely related to the theory of random sets (Nguyen, 1978); namely viewing the probability allocation m over 2^Ω as a random subset \mathcal{R} of Ω , we have $\prod(A) = P(\mathcal{R} \cap A \neq \emptyset)$ and $N(A) = P(\mathcal{R} \subseteq A)$, where P is the probability measure over 2^Ω induced by m .

A second extension consists in considering several random experiments (e.g. several urns from which to draw balls) and an unknown mixture of those that produces a probability measure P . Imprecision lies here in the lack of knowledge of the proportion of times each urn is used. If P_i is the probability measure obtained by random experiment i , $i = 1, n$, then P can be any member of the convex set \mathcal{P} generated by P_1, P_2, \dots, P_n . Uncertainty is now evaluated in terms of upper and lower probabilities

$$P^*(A) = \sup\{P(A) \mid P \in \mathcal{P}\} \quad (4)$$

$$P_*(A) = \inf\{P(A) \mid P \in \mathcal{P}\} = 1 - P^*(\bar{A}) \quad (5)$$

This type of view is advocated by Kyburg (1974) who claims that very often, we only have an incomplete knowledge of probability values, described by means of intervals of the form $[P_*(A), P^*(A)]$. Upper and lower probabilities are also generalized possibility and necessity values (Dubois and Prade, 1988a), and expected possibility and necessity, as defined by (2) and (3) correspond to special kinds of convex sets of probability measures, i.e. can be viewed as special cases of upper/lower probabilities. See Smith (1961), Giles (1982) for a betting behavior interpretation of upper and lower probabilities. $P^*(A)$ is then the least amount of money one is ready to accept, in order to bet on A with the commitment of paying \$1 if A turns out to be false. Moreover $P_*(A)$ and $P^*(\bar{A})$ are not supposed to sum to 1.

I.3 - Belief as Preference

It is not always meaningful to relate uncertainty to frequency. Indeed some types of events can be rare, unrepeatable, or not old enough so that no statistics have been obtained for these events. However people may express opinions about the likelihood of occurrence of these events. The simplest form in which such opinions can be expressed is to state that some events are more likely to occur than others. This is a purely ordinal view of uncertainty that has been proposed by De Finetti (1937). Let \mathcal{E} be a finite set of events, viewed as subsets of referential Ω . Let \succsim be a preference ordering on \mathcal{E} expressing belief. Axioms proposed by De Finetti are as follows :

A1) $A \succsim A$ (reflexivity)

A2) $A \succsim B$ and $B \succsim C$ imply $A \succsim C$ (transitivity)

A3) $A \succsim B$ or $B \succsim A$ (completeness)

A4) $\Omega \succ \emptyset$ (non-triviality)

A5) $A \succ \emptyset$ A6) if $A \cap (B \cup C) = \emptyset$ then $B \succ C$ is equivalent to $A \cup B \succ A \cup C$ (additivity)

These axioms have been studied to determine whether or not probability measures were

the only numerical representations of these orderings, where "P represents \succeq " means

$$A \succeq B \text{ if and only if } P(A) \geq P(B) \quad (6)$$

While any probability measure satisfies A1-A6, it turned out that the converse is not true. See Fine (1973) for an account of this problem, and more recently Fishburn (1986a). This negative result can be addressed by finding a substitute to (A6) that guarantee probabilistic representations. However proposed axioms were either too strong or un-natural.

Another approach is to allow for non-additive representations of belief. Namely any set-function g such that $g(\emptyset) = 0$, $g(\Omega) = 1$, that is monotonic under inclusion could serve as a numerical representation of uncertainty (Sugeno, 1977). In order to preserve the structural simplicity of probability measures, the following decomposability axiom can be adopted. There exists an operation \perp on the unit interval such that

$$\text{if } A \cap B = \emptyset \text{ then } g(A \cup B) = g(A) \perp g(B) \quad (7)$$

This kind of non-additive representations have been introduced by Dubois and Prade (1982) and Weber (1984). Similar ideas have been proposed in the literature of decision theory under the name "distorted probabilities" (Yaari, 1987; Schmeidler, 1986). An interesting question is then to find a counterpart of (A6) such that decomposable measures represent a likelihood ordering. The following axiom, a weak form of A6, has been proposed by Dubois (1986a) :

$$A7) \quad \text{if } A \cap (B \cup C) = \emptyset \text{ and } B \succeq C \text{ then } A \cup B \succeq A \cup C$$

Results in Dubois (1986a), refined by Chateauneuf (1988b), indicate that indeed the only numerical counterparts of event orderings satisfying (A1-A5)-(A7) are decomposable measures. An even simpler, but stronger axiom than A7 is obtained by dropping the empty intersection condition in A7, i.e.

$$A8) \quad \forall A, B \succeq C \text{ implies } A \cup B \succeq A \cup C$$

whose numerical representations are a special case of decomposable measures ; namely $\perp = \text{maximum}$ in (7). These set-functions are nothing but Zadeh (1978a)'s possibility measures, i.e. $\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$. The corresponding necessity measures N obtained by duality (equation (1)) are characterized by the following counterpart of A8 :

A9) $\forall A, B \succeq C$ implies $A \cap B \succeq A \cap C$

(see Dubois, 1986a) and satisfy the dual axiom $N(A \cap B) = \min(N(A), N(B))$. These set-functions are of the form (2) and (3) where the focal elements are nested ($E_1 \subseteq E_2 \subseteq \dots \subseteq E_n$ in (2) and (3)). These set functions have been first suggested as useful to model uncertainty in terms of possibility by the economist Shackle (1961). He called the quantity $N(\bar{A})$ the degree of potential surprize of A .

This connection between the purely ordinal view of uncertainty and the ones related to probabilistic models under incomplete information suggests that upper and lower probabilities can be used as degrees of belief and plausibility with no reference to any unreachable additive representation. This is the path followed by Shafer (1976) and Smets (1988) for whom degrees of subjective belief can be expressed by set-functions of the form (3), and who claim that any interpretation of (3) in terms of lower probabilities is misfounded. Fishburn (1986b) also adopts a purely ordinal view of more general classes of upper and lower probabilities, based on interval-orderings. However his axioms for Shafer belief functions are too complicated to be meaningful. Research on the determination of belief function-based orderings is going on. A natural line of research is to carry an ordering relation satisfying A1-A6) from a set of events to another one, through a multiple-valued mapping ; it comes down to using Dempster's definition based on multiple-valued mappings where the basic probabilistic assignment is changed into an ordering relation (Wong, 1991).

Recently Spohn (1988) has proposed a theory for the representation of epistemic states that bear strong similarities with possibility and necessity measures, and also try to model the idea of ordering between possible worlds in terms of belief. 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 $\{\omega\}$ such that $\kappa(\{\omega\}) = 0$. $\kappa(A)$ 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, 1991a). A probabilistic interpretation of $\kappa(A)$ has been suggested by Spohn (1990). 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 $\epsilon^{\min(m,n)}$. These remarks may lead to an interpretation of possibility and necessity measures in terms of probabilities of rare events.

I.4 - Vagueness

Part of the evidence we can collect appear under the form of natural language sentences, and not necessarily statistics in the usual sense. In traditional probability experiments, we may for instance replace the perfect observer by individuals describing results of trials not only in an incomplete manner but also in a linguistic form. For instance ill-described results of random experiments are rather common in opinion polls : people sometimes deliver linguistically expressed orders of magnitudes rather than precise numbers. Another situation is when the event whose probability is of primary concern is not perfectly defined. Namely, high level decisions are often evaluated in terms of fuzzily described goals (e.g. a given investment should produce a *significant* benefit). Hence vagueness (mainly through natural language) may pervade the description of outcomes as well as affect the notion of event, due to lexical imprecision.

Understanding a vague term in natural language assumes that one identifies the universe of concerned objects and the subclass this term refers to. In some situations this class has no definite boundaries and its representation by means of a well-defined set is but a crude approximation. The reason is that the predicates that describe the class involve gradedness. For instance "old" is an example of this kind of predicate ; even in given context and for a single locutor there is no age below which one is not old, and beyond which one is suddenly old ; one can only *become* older and older. Entering the class is a continuous process. A simple way of modelling gradedness in set membership is to use a continuous scale of membership values as proposed by Zadeh (1965), for instance the unit interval. A fuzzy set F over a referential set Ω is then defined by means of a membership function μ_F ranging on $[0,1]$. There is a lurking temptation of confusing $\mu_F(\omega)$ with a degree of probability that ω belongs to the (then well-defined) set F . This confusion is perhaps due to the choice of the unit interval for the membership scale. Anyway $\mu_F(\omega)$ is usually best interpreted in terms of proximity between ω and other elements of Ω which can be viewed as prototypes of F (and for which full membership is granted) ; then the notion of membership grade comes closer to the idea of distance rather than probability. However in some situations, membership functions have been interpreted in a probabilistic way, either as a likelihood function (Hisdal, 1988) or as the contour function of a random set (e.g. Dubois and Prade, 1989). These types of interpretation are useful when the membership function must account for the variations of a supposed threshold between membership and non-membership, according to individuals in a group:

- with the likelihood view, $\mu_F(\omega) = P(F \mid \omega)$ is the proportion of individuals who reply yes to the question "is ω an F ?". But the symbol F in $P(F \mid \omega)$ refers to the set of individuals who did reply yes (a non-fuzzy set), and $P(F \mid \omega)$ is used as a means of evaluating $\mu_F(\omega)$

- with the random set view, each individual i supplies a set E_i that is the best non-fuzzy representation of F and $\mu_F(\omega)$ is evaluated by summing the proportions of individuals i whose response E_i contains ω .

As opposed to "group membership functions", individual membership functions are not so easily liable of statistical approaches, but can be addressed by means of classical psychometrical tools (Norwich and Turksen, 1982)(Turksen, 1991).

Set-theoretical operations can of course be extended to fuzzy sets. This topic is not considered here, for the sake of brevity ; indeed there is much literature on this topic (see Dubois and Prade, 1985b for instance).

As pointed out earlier, vagueness may appear in the available evidence as well as in the description of relevant events. When the available evidence can be described in terms of a fuzzy set (or a combination thereof), this fuzzy set is viewed as an ill-defined set E of possible values of a variable, and the membership function is interpreted as a possibility distribution (Zadeh, 1978a). This is an example of incomplete information that fits the framework of Section I.2. Namely necessity and possibility measures can be defined from μ_E as follows :

$$\forall A : N(A) = \inf_{\omega \notin A} 1 - \mu_E(\omega), \quad \Pi(A) = \sup_{\omega \in A} \mu_E(\omega) \quad (8)$$

It can be checked that (8) is a special case of (2) and (3), where the set of focal elements is nested and coincide with the level cuts E_α of E ($E_\alpha = \{\omega \mid \mu_E(\omega) \geq \alpha\}$, for $\alpha \in (0,1]$) (see Dubois and Prade, 1985c). This situation corresponds to nested incomplete observations. Moreover Π is a possibility measure as induced by the preference ordering that obeys axiom A8 of the previous section.

An obvious generalization of this situation to the ideal picture of probability is the case of statistics made of fuzzy observations $\{E_i ; i = 1, n\}$. For a mathematical model of this kind of situation, see Kruse and Meyer (1987). Generalizations of possibility and necessity measures that cope with fuzzy observations have been provided by Yager (1984) and the authors (Dubois and Prade, 1985a).

Dually one may evaluate the uncertainty of ill-described events. Probability of fuzzy events has been introduced by Zadeh (1968) rather early, and has given birth to a mathematically-oriented literature focusing on the definition of so-called fuzzy σ -algebras (see Klement, 1982 for instance). Possibility and necessity measures of fuzzy events are also defined

by Zadeh (1978a, 1979) when evidence is described by means of a possibility distribution. These evaluations can be viewed as resulting from a fuzzy pattern-matching between a fuzzy event and a vague piece of evidence. The belief function of a fuzzy event (an extension of Shafer's set-function) has also been introduced by Smets (1981). Note that the main difference between the case of fuzzy events and the classical setting is that fuzzy events no longer form a Boolean algebra, but lead to weaker structures.

Lastly it should be clear that vagueness is not necessarily a drawback, a definite imperfection of the available knowledge. In the expression of queries, for instance, vague predicates may suggest the flexibility of a request rather than its lack of specification, in the same way as fuzzy sets make it possible the expression of continuous transitions from one vague category to another adjacent one, as already said.

1.5 - Indiscernibility

It is not very original to say that mathematical models sometimes fail to account for what they intend to model because they are too refined with respect to what can be perceived from the phenomenon under study. For instance the set of real numbers is clearly an example of such a mathematical construct that assumes a continuous world. Continuous probabilistic densities are idealizations of finite histograms. The finiteness of histograms is not only due to the limited number of experiments that produce them, but sometimes due to a discretization of the set of outcomes — like when establishing statistics of human sizes. In that case, the limitation of observation devices creates indiscernibility between outcomes when they are too close to each other. Another source of indiscernibility is the granularity of symbolic (e.g.natural language) descriptions, which also induces classes of indiscernible objects in computerized information systems. This type of concern has been addressed by Pawlak (1982) who starts from an equivalence relation R on a set Ω of objects, expressing indiscernibility. Two objects are indiscernible if and only if they have the same description. Due to indiscernibility, any set S of objects cannot be exactly described, since objects x can be captured only through their equivalence classes $[x]_R$. Only upper and lower images of S can be attained, namely

$$S_* = \{[x]_R \mid [x]_R \subseteq S\}$$

$$S^* = \{[x]_R \mid [x]_R \cap S \neq \emptyset\}.$$

The pair (S_*, S^*) is called a rough set. Note that histograms and digitized images are based on a similar view of indiscernibility ; pixels in digitized images are equivalence classes of indiscernible spots. Basically, in the rough set model, the set Ω of outcomes is partitioned and only subsets in the corresponding induced Boolean algebra can be used to describe the world.

Clearly rough sets are very related to the notion of inner and outer measures (Halmos, 1950), recently revived by Fagin and Halpern (1989a) in Artificial Intelligence. Namely, given a non-measurable subset S of Ω , and a probability measure defined on the quotient set Ω/R , then the probability $P(S)$ is only characterized via the interval $[P_*(S), P^*(S)]$ where $P_*(S) = P(S_*)$ and $P^*(S) = P(S^*)$, i.e. the probability of a rough set. This idea also pervades the literature on belief functions (Shafer, 1976) with the notions of coarsening and refinement of frames of discernment (the terminology here is very explicit, a frame of discernment is really of the form of a quotient set).

Although interesting, the concept of rough set alone is rather limited in order to properly account for indiscernibility. As pointed out very early by Poincaré(1902), if x is indiscernible from y , y is indiscernible from z , x and z may sometimes be told apart. In other words, indiscernibility is not transitive strictly speaking, especially on continuous universes. It suggests that indiscernibility should be represented by a weaker notion, that of a similarity relation. Similarity relations appear in different form in the theory of probabilistic metric spaces (Menger, 1951 ; Schweizer and Sklar, 1983), in cluster analysis (Jardine and Sibson, 1977), and in fuzzy set theory (Zadeh, 1971). A frame of discernment is then a set Ω equipped with a fuzzy relation R defined by its membership function $\mu_R : \Omega \times \Omega \rightarrow [0,1]$, such that $\mu_R(x,y) = 1$ means that x and y are indiscernible, and $\mu_R(x,y)$ is a degree of closeness between x and y . R satisfies the following properties

$$\begin{aligned} \mu_R(x,x) &= 1 \quad (\text{reflexivity}) \\ \mu_R(y,x) &= \mu_R(x,y) \quad (\text{symmetry}) \\ \mu_R(x,y) * \mu_R(y,z) &\leq \mu_R(x,z) \quad (*\text{-transitivity}) \end{aligned}$$

where $*$ is a conjunctive operation on $[0,1]$, i.e. coincide with a conjunction on $\{0,1\}$. The $*$ -transitivity property is weaker than the usual property of transitivity. The notion of similarity is very much akin to the one of distance, namely $1 - \mu_R$ is a distance index, which can be an ultrametric ($*$ = min) or which can satisfy the triangular inequality. The description of sets on a frame of discernment (Ω, R) , where R is a fuzzy relation, requires that extensions of equivalence classes be defined. See Valverde (1985) and Höhle (1988). Viewing the membership function of a fuzzy set in terms of similarity with respect to prototypes, any fuzzy set can be viewed as an equivalence class of a similarity relation.

The connection between rough sets and similarity relations was initiated by Fariñas del Cerro and Prade (1986) and completed in Dubois and Prade (1990b). Despite the close relationship between similarity relations and fuzzy sets, it is clear that they address different issues : indiscernibility of objects versus graduality in set-membership. For instance the size of

the pixels in a digitized image is not the same parameter as the number of levels of grey. However the lack of transitivity of indiscernibility relations is clearly related to the graduality of this notion.

II - Conditioning, Independence, Expectations and the Combination of Information

In the preceding sections, we have shown various alterations to the ideal picture of probability that led to non-additive representations of uncertainty and to gradedness in set-membership. A natural question that arises is whether well-established concepts of probability theory, especially conditioning, independence, and expectation carry over to the non-additive settings. Moreover it can be claimed that the emergence of non-additive probabilities has shed light on the problem of combination of evidence, a problem that worried scholars in the XVII^e century and that had remained in the shade until recently (Shafer, 1986).

II.1 - Conditioning

For simplicity, let us restrict ourselves to expected possibility and necessity functions of the form (2), (3), i.e. the set-functions introduced by Dempster (1967) and used by Shafer(1976) in the theory of belief functions. Conditioning, and more specifically conditional probability can be extended in more than one way ; if $\Pi(A)$ and $N(A)$ are understood as probability bounds, then a first approach is to start with the corresponding family \mathcal{P} , and to consider the family $\mathcal{P}_B = \{P(\cdot | B) | P \in \mathcal{P}\}$ obtained by conditioning each member of the family. This way we obtain upper and lower conditional probabilities $P^*(\cdot | B)$ and $P_*(\cdot | B)$ by (4) and (5) applied to \mathcal{P}_B . This approach to conditioning has been advocated by Kyburg (1987), Fagin and Halpern (1988) among others. Interestingly $P_*(\cdot | B)$ is still a belief function in the sense of Shafer (Fagin and Halpern, 1989b ; Jaffray, 1990).

Another view of conditioning, more popular in the theory of belief functions, is to directly extend the formal definition to expected possibility functions :

$$\Pi(A | B) = \frac{\Pi(A \cap B)}{\Pi(B)} \quad (9)$$

and $N(A | B) = 1 - \Pi(\bar{A} | B)$. This is Dempster rule of conditioning (Shafer, 1976). This form can be justified in the setting of Cox (1946)'s axiomatic justification of probability theory (Dubois and Prade, 1990a). Namely given a set-function g defined on a Boolean algebra of

events, the conditional set-function $g(A|B)$ is supposed to satisfy the following property

$$\forall A, g(A \cap B) = f(g(A|B), g(B)) \quad (10)$$

which defines the uncertainty of $A \cap B$ as a function f of the uncertainty of B and the uncertainty of A given that B occurs. When f is strictly increasing in both places, f is isomorphic to the product, which justifies (9). Cox (1946) has proved that if moreover $g(A)$ entirely determines $g(\bar{A})$, $\forall A$, then g should be a probability measure ; however this assumption cannot be made in the presence of incomplete information. Interpreting possibility measures in the consonant case using Spohn (1988)'s conditional functions (see section I.3), that is $\prod(A) = k^{-\kappa(A)}$, one is led from (9) to define $\kappa(A|B) = \kappa(A \cap B) - \kappa(B)$, which is exactly Spohn's proposal for conditioning. The same author has moreover proved that this conditioning rule is in accordance with usual conditional probabilities when the $\kappa(B)$'s are related to infinitesimal probabilities as recalled in section I.3.

Clearly a third alternative is to apply (9) to expected necessity measures ; this is the so-called geometric rule of conditioning (Suppes and Zanotti, 1977). Note that upper and lower conditional probabilities $P^*(\cdot|B)$ and $P_*(\cdot|B)$ introduced above do have a closed form for quite a large family of upper and lower probability functions, namely (De Campos et al., 1990)

$$P^*(A|B) = \frac{\prod(A \cap B)}{\prod(A \cap B) + N(\bar{A} \cap B)} ; P_*(A|B) = \frac{N(A \cap B)}{N(A \cap B) + \prod(\bar{A} \cap B)} \quad (11).$$

(11), (9) and its companion in terms of expected necessity functions are generalizations of conditional probabilities. Clearly, (11) makes sense only if \prod and N are interpreted as probability bounds otherwise, if the uncertainty function stems from preference ordering only (10) makes sense. See Weber (1988) for conditional decomposable measures (see (7)) in the spirit of (10).

For expected possibility and necessity functions, it is interesting to define conditioning in terms of the underlying focal elements E_1, \dots, E_n , and the basic probabilities $m(E_i)$ $i = 1, n$. This is very simple for Dempster rule (9) since the basic probabilities after conditioning become

$$m(A|B) = \frac{1}{k} \sum \{m(E_i) | B \cap E_i = A\}, \forall A \quad (12)$$

where k is a normalization factor whose purpose is to preserve

$$\sum_A m(A|B) = 1.$$

It means that all masses $m(E_i)$ such that $E_i \cap B \neq \emptyset$ move to $E_i \cap B$ and the remaining masses are proportionally added to the moved masses. In the case of the geometric rule, only masses $m(E_i)$ with $E_i \subseteq B$ are kept and then re-normalized.

For the upper and lower conditioning rule (11), it has been proved by De Campos et al. (1990) that $P^*(\cdot | B)$ and $P_*(\cdot | B)$ can be determined in two steps :

- Given a vector ν of quantities that describes, for each subset $C \subseteq \Omega$, the probability $P(B | C)$ that some object for which we only know that it belongs to C , does belong to B , we can compute a conditional probability mass :

$$m_\nu(A | B) = \frac{1}{k} \sum_{E_i} \{m(E_i) \cdot P(B | E_i) | B \cap E_i = A\}, \forall A \quad (13)$$

But ν is not supposed to be known in the upper and lower probability setting.

- Then $P_*(A | B) = \inf_\nu N_\nu(A | B)$, $P^*(A | B) = \sup_\nu \prod_\nu(A | B)$, where N_ν and \prod_ν are expected possibility and necessity functions based on $m_\nu(\cdot | B)$. Interestingly, $P_*(A | B)$ and $P^*(A | B)$ are still expected necessity and possibility functions that obey the laws of belief function theory (Fagin and Halpern, 1989b; Jaffray, 1990).

Note that although ν is not known, some obvious constraints must be respected, namely $P(B | C) = 0$ if $B \cap C = \emptyset$ and $P(B | C) = 1$ if $C \subseteq B$. $P(B | C)$ is unknown only if B and C are logically independent. This result enables us to see that Dempster rule of conditioning implicitly assumes that $\forall C, B \cap C \neq \emptyset$ implies $P(B | C) = 1$, while the geometric rule of conditioning assumes that $P(B | C) = 0$ unless $C \subseteq B$. This explains why the ranges $[P_*(A | B), P^*(A | B)]$ are wider than the ranges pertaining to other conditioning rules (Kyburg, 1987).

This analysis provides a semantic interpretation to the various conditioning notions, and explains why, in the presence of incomplete information, it is natural to find more than one conditioning rule in this situation :

- upper and lower conditioning corresponds to changing the reference class of objects. It provides the answer to the question : given what is known of objects in Ω , what can be said about objects lying in B ? The problem is that, due to incomplete information (e.g. set-valued statistics), it may be impossible to decide about membership to B for some of the concerned objects ;
- Dempster rule of conditioning correspond to an updating process, i.e. a new piece of evidence

arrives, namely that the object to be located does belong to B , hence each focal element E_i can be refined into $E_i \cap B$, and $m(E_i)$ can move to this smaller set.

Hence the first type of conditioning is a sort of zooming or focusing process and does not presupposes the arrival of a new piece of information, while the second type does correspond to a revision process. In probability theory the two notions (focusing and updating) coincide and are expressed by conditional probability ; this is not surprising since $m(C) > 0$ only if C is a singleton, and $P(B | C) \in \{0,1\}$ in (13). This fact points out some limitations of the expressive power of probability theory in the presence of incomplete information.

Of course, the next intriguing step is the extension of Bayes theorem and Bayesian inference to non-additive probability. Conditional functions based on (10) very neatly lend themselves to an extension of Bayes theorem ; this is not so true for upper and lower conditional probabilities (11). But Bayesian inference deals with updating processes while (11) does not, so that recovering an analog to Bayes theorem from (11) is perhaps meaningless.

Conditional probability can be extended in another direction, by changing events into fuzzy events. See M.A. Gil (1988) for a survey on this trend and its applications to statistics.

II.2 - Independence

The question of extending the concept of independence beyond the probabilistic setting has not been considered so often. If a non-additive set-function is used to describe bounds on the value of an unknown probability, the classical notion of independence can be applied to the unknown probability. What should be done when a non-additive uncertainty function is the representation of a preference relation is not so clear, as long as no frequency-like view is allowed. One starting point may be to restrict to purely formal arguments, and solve the following problem : let g be an uncertainty measure of some kind, what is the set of possible operations $*$ such that $g(A \cap B)$ can be expressed in the form $g(A) * g(B)$. The monotonicity of g with respect to set-inclusion obviously forces $g(A \cap B) \leq \min(g(A), g(B))$. Moreover $g(A \cap B) \geq \max(0, g(A) + g(B) - 1)$ holds for set-functions that coincide with upper and lower probabilities. When g is a probability measure, a unique family of operations $*$ exists (Dubois, 1986b).

Besides, the presence of incomplete information leads to distinguish between two kinds of assumptions in the probabilistic framework, when solving the problem of finding $P(A \cap B)$ given the knowledge of $P(A) = \alpha$ and $P(B) = \beta$

- assume that A and B are independent. Then $P(A \cap B) = P(A) \cdot P(B) = \alpha \cdot \beta$
- assume that the dependency links between A and B are unknown. This leads to the inequalities $\min(\alpha, \beta) \geq P(A \cap B) \geq \max(0, \alpha + \beta - 1)$ and to upper/lower probability settings.

The first attitude has been justified through the principle of maximal entropy. The second attitude can be advocated by the principle of minimum specificity (Dubois and Prade, 1986a) which tries to preserve incompleteness for the sake of being faithful to the amount of available information. Possibility theory (Zadeh, 1978a) which uses set-functions Π and N of the form (8), based on fuzzy sets heavily relies on this principle. Especially it justifies the choice of the minimum operation when combining possibility distributions issued from several sources, or pertaining to so-called non-interactive variables (Zadeh, 1975). In this framework the minimum specificity principle leads to choose the least restrictive possibility distribution compatible with what is known, that is, the one which allocates the highest degrees of possibility. The price paid when using either principles is quite different : assuming independence may lead to wrong conclusions; assuming a lack of dependence knowledge may produce too imprecise or even vacuous conclusions.

II.3 - Expectation

Another exciting topic is expectation. Contrary to independence, it is the subject of a growing literature from mathematicians, and more recently by people in decision theory. There are two kinds of expectation measures in the field of non-additive probability : upper and lower expectations, and what can be termed "distorted" expectations. The first type of expectation heavily borrows from early works by Choquet (1953), and best fits uncertainty representations that refer to upper and lower probabilities, or to possibility/necessity pairs. The simplest way of obtaining an upper expectation is to consider a family \mathcal{P} of probability measures, select a member P of this set, compute a classical expectation \mathcal{E} from P and obtain $\mathcal{E}^* = \sup\{\mathcal{E} \mid P \in \mathcal{P}\}$; similarly, the lower expectation $\mathcal{E}_* = \inf\{\mathcal{E} \mid P \in \mathcal{P}\}$. Other ways of expressing upper and lower expectations exist. For instance, assume $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, and a pair of inclusion-monotonic set-functions (N, Π) , such that $N \leq \Pi$, is given as well as a function f from Ω to \mathbb{R} that assigns numerical values to the ω_i . Then it is easy to associate upper and lower distribution functions F_N and F_Π to N and Π , defining

$$F_N(u) = N(\{\omega_i \mid f(\omega_i) \leq u\}) \quad ; \quad F_\Pi(u) = \Pi(\{\omega_i \mid f(\omega_i) \leq u\}) \quad (14)$$

and then letting

$$\mathcal{E}^*(f) = \int_{-\infty}^{+\infty} u \cdot dF_N(u), \mathcal{E}^*(f) = \int_{-\infty}^{+\infty} u \cdot dF_{\prod}(u)$$

For instance these definitions can be used with expected possibility and necessity functions, and are used by Smets (1981) to compute the degrees of belief and plausibility measures of fuzzy events, and by the authors to compute mean values of fuzzy numbers (Dubois and Prade, 1987a). These notions also have applications, in decision theory, to the computation of utility functions in the face of uncertainty, so as to cope with paradoxes of classical expected utility ; see Yaari (1987), Gilboa (1987), Chateauneuf (1988a), Schmeidler (1986), Wakker (1989), among others.

Another view of expectation relates to decomposable measures (defined by (7)) where additivity is changed into another form of compositionality. The idea is to distort the classical expectation likewise. A first example of this kind of integral was suggested by Sugeno (1977) under the name "fuzzy integral". Let g be a monotonic set-function over Ω , and $A_\alpha = \{\omega \mid f(\omega) \geq \alpha\}$ for $\alpha \in [0,1]$, then

$$\tilde{\mathcal{E}}(f) = \sup_{\alpha} \min(\alpha, g(A_\alpha)) \quad (15)$$

This kind of expectation makes sense for numerical functions f that range in the unit interval and more generally, the same range as the set-function g . The analogy between (14) and (15) is striking. (15) can be interpreted as a median rather than an expectation and especially makes sense when g is a max-decomposable set-function, i.e. a possibility measure in the sense of Zadeh (1978a). Namely if \prod derives from the membership function μ_E as in (8), then (15) simplifies into

$$\tilde{\mathcal{E}}(f) = \max_{\omega \in \Omega} \min(f(\omega), \mu_E(\omega)) \quad (16)$$

and can be interpreted as the possibility degree of a fuzzy event F with membership function f ; indeed (16) evaluates the consistency between F and E , in accordance with the meaning of possibility in Section I. This kind of expectation has been extended to other types of decomposable measures by Weber (1984) and Murofushi and Sugeno (1989).

II.4 - The Combination of Evidence

Conditioning is a tool that, in probability theory, captures the accumulation of evidence appearing under the form of observed events, by updating an a priori probability distribution. All cases for information combination cannot be cast in this setting in a natural way. Shafer (1986)

points out the example of combining unreliable testimonies as a problem that was considered quite early in the history of probability, and that was given a non-Bayesian solution. Namely, if two witnesses claim that some event B has occurred, that witness 1 has probability p_1 of being reliable, while the reliability of witness 2 is p_2 , then, assuming that these are independent pieces of evidence, the confidence in the occurrence of B can be measured by $p_1 + p_2 - p_1p_2$ (Hooper's rule). This type of problem has been also recently considered by Swedish researchers (Gärdenfors et al., 1983) and is the topic of Shafer's theory of evidence. There are two differences between the problem of combining testimonies and the kind of situations addressed by Bayesian inference : in the latter, one of the elements to be combined is not uncertain ; moreover the two pieces of information do not play the same role.

In fact, we can distinguish between two types of combination modes :

- the combination of evidence issued from concurrent sources. It corresponds to a symmetric mode of combination ;
- the updating of an epistemic state of knowledge on the basis of some observation. It corresponds to a dissymmetric mode of combination since updating often aims at keeping as much as possible of the previous state of knowledge while taking into account the observation.

Of course the two modes of combination can be simultaneously employed when updating a state of knowledge on the basis of concurrent pieces of evidence. The Bayesian approach to the combination of evidence always assumes that some a priori knowledge exists, so that only updating makes sense. This is not the case for other types of uncertainty models, which allow to accumulate evidence from scratch. A comparison of Bayesian updating from concurrent observation and the symmetric combination solution in the setting of belief functions is given by Shafer (1986).

Symmetric rules

In the theory of belief functions, the main symmetric rule of combination is Dempster rule (Dempster, 1967). Namely given two belief functions on Ω , defined via two probability allocations m_1 and m_2 over $2^\Omega - \{\emptyset\}$, the set of non-empty subsets of Ω , the result of the combination is represented by a belief function whose probability allocation is defined by

$$\forall C \subseteq \Omega, m(C) = \frac{1}{k} \sum_{C=A \cap B} m_1(A) \cdot m_2(B) \quad (17)$$

where k is a normalization factor equal to $1 - \sum_{A \cap B = \emptyset} m_1(A) \cdot m_2(B)$. This rule works in two

steps : an intersection of the random sets \mathcal{R}_1 and \mathcal{R}_2 , supposedly independent ; a conditioning step that updates the result of the random set intersection on the fact that the intersection is not empty. Thus interpreted on 2^Ω , we can write $m(C) = P(\mathcal{R}_1 \cap \mathcal{R}_2 = C \mid \mathcal{R}_1 \cap \mathcal{R}_2 \neq \emptyset)$.

As an illustration, let us consider the concurrent testimony case, where two independent witnesses report that $B \subseteq \Omega$ occurred. The reliability p_1 of witness 1, is interpreted as follows : p_1 is the probability that witness 1 is faithful, and $1 - p_1$ the probability that he is erratic. Namely $1 - p_1$ accounts for the case when it is impossible to conclude anything from the report of the witness. This is modeled by the random set $m_1(B) = p_1$, $m_2(\Omega) = 1 - p_1$, and similarly for witness 2 whose reliability is p_2 . Applying (17) leads to $m(B) = p_1 + p_2 - p_1 p_2$, $m(\Omega) = (1 - p_1)(1 - p_2)$ ($k = 1$ in that case) ; this is Hooper's rule (already proposed at the end of the XVIIth century, see (Shafer, 1986)).

Counterparts of Dempster rule in possibility theory exist, changing random set intersection into fuzzy set intersection, and adapting the normalization factor (see Dubois and Prade, 1988d).

Note that the probability allocation m_1 qualifies the reliability of witness 1 as a transmitter of an observation he made. $m_1(B)$ is the probability that witness 1 observed B given that he says he observed it. The summation in (17) then presupposes that the reliability of both witness as observers is total, and that both witnesses observed the same phenomenon. For instance if witness 1 observed event B and witness 2 event B' , this implies $B \cap B' \neq \emptyset$, and that event $B \cap B'$ occurred. The product $m_1(B) \cdot m_2(B')$ is simply the joint probability of witness 1 observing B and witness 2 observing B' ; the normalization factor updates this joint probability by the fact, taken as a piece of evidence, that both witnesses are reliable observers (but not necessarily transmitters) and did observe the same phenomenon.

When the witnesses are no longer reliable observers, then a weaker assumption is that one of them is a good observer. So, when witness 1 observes B and witness 2 observes B' from the same phenomenon, the only meaningful conclusion is that the truth lies in $B \cup B'$. It leads to the disjunctive counterpart of Dempster rule which writes

$$m(C) = \sum_{C=B \cup B'} m_1(B) \cdot m_2(B') \quad (18)$$

This rule has been first used by Smets (1978) and advocated by the authors (Dubois and Prade, 1986b) in a set-theoretic perspective. Applied to the case of Hooper rule, it is easy to check that combining testimonies from unreliable observers leads to believe event B with strength

$p_1 p_2 < \min(p_1, p_2)$. The negative reinforcement effect can be explained as follows : one reliable observer is better than two when one of these two is not reliable, if we do not know which one is reliable.

In a Bayesian view of this problem, p_1 becomes the probability that witness 1 is truthful, while $1 - p_1$ is the probability that witness 1 lies. The witness is never assumed to be careless nor a bad observer. Moreover it is assumed that some a priori probability $P(B)$ of B is available, which does not account for the witnesses view. Using Bayes theorem, it is possible to derive the probability $P(B \mid \text{witnesses report } B) = P'(B)$ under the form

$$P'(B) = \frac{P(B)p_1 p_2}{P(B)p_1 p_2 + (1 - P(B))(1 - p_1)(1 - p_2)} \quad (19)$$

If $P(B) = 1/2$ (the Bayesian interpretation of the absence of a priori information) one gets $P'(B) = p_1 p_2 / (p_1 p_2 + (1 - p_1)(1 - p_2))$. It is easy to see that this is exactly what would be obtained by Dempster rule (17), if $1 - p_1$ is interpreted as $m_1(\bar{B})$ instead of $m_1(\Omega)$. As Shafer (1986) points out, the difference between the Bayesian approach to the combination problem and the belief function approach is that the former demands a full probability model (e.g. the existence of prior knowledge, the probability that a witness is mendacious, etc.).

Contrasting with the Bayesian rule (19), there exists another combination rule in probability theory that is symmetric with regard to the sources of information, and enjoys an idempotence property not satisfied by (17) and (18). Namely

$$P''(B) = \frac{P_1 + P_2}{2} \quad (20)$$

This rule can be justified by considering the witnesses as equivalent random sources where p_1 is the proportion of times witness 1 would be truthful, and $1 - p_1$ the proportion of times he would be lying. The combination rule just sums up the two experiments. Lehrer and Wagner (1981) have justified this kind of rule from purely axiomatic arguments. The arithmetic mean rule is the only one such that $P(A) = f(P_1(A), P_2(A))$, $\forall A \subseteq \Omega$, f is symmetric, and $f(0,0) = 0$, $f(1,1) = 1$. Dropping the symmetry condition one gets all convex mixtures of probabilities. Consensus rules such as (20) make sense when modeling voting procedures, i.e. p_1 reflects the probability that individual 1 would choose alternative B , and $P''(B)$ reflects the opinion of a group. This is clearly not the same situation as with the witness problem. Rule (20) applies to belief functions as well (Wagner, 1989).

Updating rules

The updating problem contrasts with the one of symmetric combination in the desire to keep as close as possible to the prior knowledge while integrating the new piece of evidence as an acknowledged fact. Interestingly, updating has been formalized in the case of uncertain prior knowledge in the setting of probability theory for a long time. The case of prior knowledge modelled by means of a set of formulas in propositional logic has been only recently studied by Gärdenfors (1988).

In the classical probabilistic case, given an a priori probability P and a sure observation B , the result of updating P by B is defined by $P(\cdot | B)$. Observation B is acknowledged since $P(B | B) = 1$. Moreover $P(\cdot | B)$ does obey the principle of minimal change which updating presupposes, since $P(\cdot | B)$ is the probability measure that is the closest to P among the measures P' that satisfy $P'(B) = 1$, in the sense of relative entropy (Williams, 1980).

This approach can be extended to uncertain observations. Bayesian updating in the face of uncertain evidence has been considered by Jeffrey (1983). Namely observing event B with probability p , leads to update a prior probability P into a posterior probability P' defined by

$$\forall A, P'(A) = p \cdot P(A | B) + (1 - p) \cdot P(A | \bar{B}) \quad (21)$$

This updating rule can also be justified via maximum entropy arguments. (21) is actually a convex mixture of the two possible opposite updated probabilities. As seen earlier, the natural updating rule in the theory of belief functions is Dempster rule of conditioning (9), where a piece of evidence B is used to refine the focal elements of the belief function representing the prior knowledge. The natural extension of Jeffrey's rule to belief functions is not Dempster rule of combination (17) since the latter is symmetric. Namely let the prior knowledge be defined in terms of an expected possibility function Π and the uncertain observation is obtained as a basic probability assignment such that A_i is observed with probability m_i , $i = 1, n$. The updating rule consists in updating Π on the basis of observations B_1, \dots, B_n separately, and to compute the weighted average of the result, i.e.

$$\forall C, \Pi'(C) = \sum_{i=1, n} m_i \cdot \Pi(C | B_i) \quad (22)$$

When Π is a probability measure, (22) coincides with Jeffrey's rule. A major difference between (17) and (22), is that Dempster rule of combination is mathematically defined whenever $\exists A, B$, $m_1(A) > 0$, $m_2(B) > 0$ and $A \cap B \neq \emptyset$ (i.e. $k \neq 0$), while (22) requires that $\forall i, \exists A, m(A) > 0$ and $B_i \cap A \neq \emptyset$, i.e. $\forall i, \Pi(B_i) > 0$. The justification of (22) in terms of minimizing some information distance is an open problem, because although information measures exist in

evidence theory (Klir, 1987), relative information measures for belief functions remain little developed to-date.

The updating rule (22) cannot be used as such for possibility measures because they are not closed under convex mixtures. However a counterpart of Jeffrey's rule has been proposed elsewhere (Dubois and Prade, 1991a).

III - Non-Standard Logics of Uncertainty and Vagueness

As said at the beginning of the paper the question of devising automated reasoning machineries on computers has led to an increasing concern of computer scientists for logics, formerly a field cherished by mathematicians and philosophers mainly. The interest in logic stems from its ability to perform domain independent reasoning tasks, owing to the separation between syntax and semantics. While general theorem-proving techniques for classical logic used as a knowledge representation language have been widely investigated, a challenging problem is to extend automated reasoning research to the case of incomplete, uncertain or vague knowledge bases.

The starting point of this type of investigations seems to be the handling of "certainty factors" in production rule systems, also called "expert systems" (Buchanan and Shortliffe, 1984). Many rules for combination and propagation of certainty factors have been proposed ; however no clear semantics of these rules has been proposed in this literature. A proper treatment of uncertainty in knowledge-based systems requires that more fundamental investigation on the links between uncertainty, vagueness and logic be carried out. This section gives an outline of the current state of the art. A first distinction is drawn between logics of vagueness where a certain view of intermediary truth-values exist and logics of uncertainty that express a lack of knowledge about the truth or falsity of statements. Logics of uncertainty are shown to be manifold, and various kinds of uncertainty logics are characterized. The specificity of logics of vagueness is then discussed, with emphasis on the problem of truth-functionality.

III.1 - Degree of Truth Versus Degree of Uncertainty

It seems very important, both at a conceptual and at a practical level, to distinguish between degree of truth and degree of uncertainty attached to a proposition (Dubois and Prade, 1988b). In this paper we shall not be concerned by fundamental questions about the meaning of truth. A practical view of truth is adopted. A truth-assignment function is defined here as a mapping t from a set of well-formed expressions in a logic to the unit interval, or a subset of it

taken as a set of degrees of truth, which is compositional with respect to logical connectives (this is called truth-functionality, i.e. $t(\neg p) = \varphi(t(p))$ and for any binary connective $*$, $t(p * q) = \psi(t(p), t(q))$). Multiple-valued logics proposed by Lukasiewicz, Kleene, Bochvar, etc. are based on such kinds of degrees of truth. Moreover, since logics are considered in an information systems perspective, the degree of truth of a sentence is understood as an evaluation of the conformity of the meaning of this sentence with a description of reality that reflects available evidence.

Two cases will be distinguished according to whether the available evidence is complete or not. If the available evidence is complete, any proposition can be said to be true or false, as long as the sentences that encodes this proposition obeys the rules of classical logic. The notion of partial truth (between true and false) can be naturally interpreted in relation to the vagueness of a proposition to evaluate, due to the presence of fuzzy predicates or quantifiers. For instance, the formula *tall(John)* expressing 'John is tall' may be partially true (say at level 0.8), given that the available knowledge about John is *complete*, i.e. if John is actually 1.75 meter tall, and if the degree of membership of 1.75 in the set of tall people sizes in a given context is indeed 0.8. It means that *tall* does not completely apply to John, although it is a rather relevant qualifier.

As already said, this view of truth considers degrees of truth as resulting from a matching process between a statement and what is known about reality (here what is known about John's size) — but not reality itself. This view corresponds to what Zadeh (1982) calls "test-score semantics". When the available information is incomplete, the calculation of truth in a precise way is generally not possible, and the use of degrees of uncertainty accounts for the imperfect knowledge of truth. Especially when the proposition to evaluate is crisp (i.e. non-fuzzy), it may not be possible to assert that the proposition is true nor that it is false. Only the degree of uncertainty that the proposition is true and the degree of uncertainty that the proposition is false can be evaluated. Clearly these degrees do not express partial or intermediary truth, but propensity to be true, to be false. The properties of the degrees of uncertainty depend upon the model that embodies the available information. In the presence of statistical information about John's size, the degree of uncertainty that "John is more than 1.70 m tall" will be a degree of probability. In the presence of vague knowledge (e.g. knowing that John is *moderately tall* defined by a possibility distribution based on the membership function of *moderately tall*), the uncertainty that "John is more than 1.70 m tall" will be expressed in terms of possibility and necessity measures.

Clearly the two situations (evaluating vague statements under complete information versus evaluating non-vague statements under incomplete information) are quite different. Yet, they can be represented in a similar way by means of a sentence in logic to which one or two

numbers are attached — hence a confusion that pervades the literature on expert systems. Actually, degrees of truth and degrees of uncertainty do not behave similarly with respect to disjunction, conjunction and negation of logical sentences. Degrees of uncertainty attached to logical sentences in a Boolean algebra are not compositional (see e.g. Dubois and Prade, 1988b ; Weston, 1987), i.e. the degree of uncertainty of a conjunction or a disjunction is generally not a function of the degrees of uncertainty of the elementary sentences involved in the conjunction or a disjunction. At least, compositionality cannot hold for all connectives of logic and all propositions in a Boolean algebra. For instance, possibility measures are compositional for disjunctions only, and probability measures for negation only. This is because Boolean algebras cannot be made isomorphic to the unit interval.

By contrast, degrees of truth of vague propositions can be truth-functional, and many-valued logics serve as a framework for reasoning with vague propositions (e.g. Pavelka, 1979). This is because, generally, algebras of vague propositions are no longer Boolean, and devising isomorphisms between these algebras and the unit interval makes sense. For instance using \max , \min , and $1 - (\cdot)$ to express disjunction, conjunction and negation of fuzzy propositions equips closed sets of such propositions with a distributive lattice structure that is compatible with the unit interval. Sometimes, arguments against fuzzy set theory rely on the impossibility of compositionality (e.g. Weston, 1987). Usually these arguments are based on the wrong assumption that the algebra of propositions to be evaluated is Boolean.

The confusion between graded "truth" and uncertainty seems to be at the root of paradoxes of early 3-valued logic such as Kleene's. It is not self-consistent to add a third truth value ν to $\{0,1\}$, that expresses "unknown", to admit that sentences like $p \vee \neg p$ are tautologies, and to preserve truth-functionality. This contradiction has been used to refute many-valued logics, since then $t(p \vee \neg p) = 1$ regardless of $t(p)$ and $t(\neg p)$. The only way out of this dilemma is either to admit that the third "truth-value" is a degree of uncertainty that stands for the set $\{0,1\}$ itself, or to admit that whenever $t(p) = \nu$, p cannot be viewed as an element in a Boolean algebra, e.g. it is a fuzzy proposition whose conformity with a complete description of reality can be a matter of degree. The first choice leads to so-called possibilistic logic, in which the truth-value $\nu = \{0,1\}$, is interpreted as a vacuous possibility distribution on $\{0,1\}$ such that $\Pi(0) = \Pi(1) = 1$. Possibility measures being truth-functional for disjunction only, we have $\Pi(\neg p \vee p) = \max(\Pi(\neg p), \Pi(p)) = 1$ which solves the paradox : indeed we may have $\Pi(\neg p) = \Pi(p) = 1$ (i.e. $t(p) = \nu$), and $\Pi(\neg p \wedge p) = 0$ (i.e. $p \wedge \neg p$ is a contradiction), since Π is not truth-functional for conjunction ($\Pi(\neg p \wedge p) \leq \min(\Pi(p), \Pi(\neg p))$ only). The second choice leads to the fuzzy set-interpretation of multiple-valued logics after Goguen (1969) and Pavelka (1979).

As a consequence of the above view of truth-evaluation, it does not make sense to assign

degrees of truth to crisp propositions in classical logic. An expression like " $x = 5$ " is either true or false. Yet, it is not meaningless to say "It is *almost true* that $x = 5$ " if this statement is interpreted by " x is *approximately* equal to 5". This interpretation points out a dispositional use of crisp statements that need to be fuzzified before their actual meaning can be laid bare. This can be done by equipping the domain of the statements with a fuzzy proximity relation R such that " $x = a$ " actually means " $x \in R(a)$ " where $R(a)$ is the fuzzy set of elements close to a (e.g. Prade, 1985, p. 269). This idea is close to Weston's (1987) concept of approximate truth as reflecting a *distance* between a statement and "ideal truth". And indeed membership grades $\mu_F(\omega)$ are sometimes viewed as based on the distance between ω and a prototype element of the fuzzy set F , especially in pattern classification problems (Bezdek, 1981). The relation R is implicit in some crisp statements and expresses indiscernibility between values of the underlying variables, and the degree of truth of " $x = 5$ " given that $x = 4$, is calculated as $\mu_{R(5)}(4)$. This proposal is the basic idea underlying Ruspini (1991)'s semantics of fuzzy logic. This dispositional use of crisp statements, underlying an indiscernibility relation, contrasts with Zadeh's (1987) frequentist approach to dispositions. He considers interpretations of "snow is white" as "usually, snow is white" while here, we suggest that it may also mean "snow is sort of white".

Note that the knowledge of a truth value attached to a vague proposition should not allow a precise piece of information to be recovered. Namely if it is known that John is 1.75 m, then the truth value of the formula $tall(John)$ is, say $\mu_{tall}(1.75) = 0.8$ once μ_{tall} has been identified. But if it is known that the degree of truth of $tall(John)$ is 0.8, it would not make sense to infer that John is 1.75 m tall precisely. This is one of the classical objections against fuzzy sets. This paradox is solved once it is acknowledged that the scale $[0,1]$ for truth-values is a convention, and that the use of this scale is dictated by convenience and simplicity ; the truth-value scale is far too refined for what it is meant to express ; only a finite number of truth values are really meaningful. One way of capturing this fact is to equip the unit interval with a fuzzy proximity relation R such that when the truth value is t_0 it really means that it is in the neighborhood of t_0 , i.e. $\tau = R(t_0)$. Given that it is *rather true* ($\tau = R(0.8)$) that John is tall, one may infer that John is "about 1.75 m" tall, where "about 1.75 m" has membership function $\mu = \mu_{R(0.8)} \circ \mu_{tall}$, following Zadeh (1979b) (where \circ denotes the composition of functions and $\mu_{R(0.8)}$ is the membership function of the fuzzy set of values close to 0.8).

On the whole, membership grades and possibility/necessity degrees lead to two kinds of logics, fuzzy logics that model vague statements, and uncertainty logics that incompletely describe the truth or falsity of crisp statements. These situations can be combined, as in truth-evaluation of vague statements in the presence of incomplete information. In this latter case the evaluation produces sets, fuzzy sets, random sets of truth values (according to the nature of available information) which can be summarized by means of degrees of uncertainty. For

instance Zadeh's (1968) probability of a fuzzy event can be viewed as an expected truth-value. Zadeh's possibility of a fuzzy event is "a fuzzy expected truth-value" in the sense of Sugeno's integral (Sugeno, 1977). Smets' (1981) degree of belief in a fuzzy event is a lower expectation (in the sense of Dempster (1967)) of truth.

Based on the distinction between degrees of truth and degrees of uncertainty, the various numerical approaches to automated reasoning split into two quite distinct types of logics : logics of uncertainty and logics of vagueness.

III.2 - Logics of Uncertainty

The term "logics of uncertainty" does not seem to be very usual in the literature of logic. For instance, the article on non-standard logics in the Encyclopedia Universalis does not refer to this terminology, while there is a section on fuzzy logics (Dubucs, 1989). Yet, the last 10 years have witnessed a growing interest of the Artificial Intelligence community for uncertainty handling in automated reasoning systems.

III.2.1 - Probabilistic Logics

The idea that classical logic may not be enough to model the way humans reason, or even ought to reason, and that probability theory should play a role in that matter is not new. This concern appears in G. Boole's works (1854) on the laws of thought. Since then there have been several attempts to let probability degrees into formal logic systems. Reichenbach (1949) tried to marry logic and frequentist probability so as to obtain a (non truth-functional) many-valued logic system. Carnap (1950) put forward an objectivist view of probability in logic based on the idea of confirmation of an hypothesis h based on evidence e , both expressed by sentences in first-order logic. The degree of confirmation $C(h,e)$ is interpreted as the conditional probability $P(h | e)$ derived from a probability distribution over the set of interpretations, the idea being that $C(h | e)$ is the proportion of worlds that make h true among those which evidence points out as possible. The confirmation function is supposed to be almost determined by the language in which h and e are described (up to a positive real-valued parameter).

Along that line, a view of probabilistic logic based on the knowledge of a probability measure on a set of interpretations has been developed by Løf (1963), Fenstad (1967) for instance. Since the probability of any interpretation is known, the probability of any sentence can be computed ; but this probability distribution can stem from frequencies of observations or subjective probabilities. Along this line of thought is the recent literature on probabilistic reasoning with Bayesian networks (Pearl, 1988) where the probability distribution on the set of

interpretations is described as a product of conditional probabilities that form an acyclic graph. The use of the graph representation seems to be more efficient than the explicit handling of a probability distribution over a set of interpretations.

Contrastedly, another view of probabilistic logic starts from a given set of sentences to which probability degrees are attached. They define constraints on a family of probability distributions over the set of interpretations. Usually only bounds on the probability of other sentences can be computed. This problem seems to have been first considered by De Finetti (1937) as follows : given a set of events E_1, \dots, E_n with known probabilities what can be said of the probability of another event E ? This problem has been further studied by Adams and Levine (1975) and Nilsson (1986). Clearly this view of probabilistic logic is in accordance with upper and lower probability systems, because it is clearly an example of incomplete knowledge about probabilities. This view of probabilistic logic enables conditional statements involving exceptions to be modelled. For instance, a statement such as "most students are adults" can be modeled as $P(\text{Adult} \mid \text{Stu}) \geq \alpha$ where α is close to 1. Hence, in order to be fully expressive, probabilistic logics must allow for sentences of that kind, which cannot be expressed by means of a sentence in first-order logic and a probability attached to it. Especially $P(\text{Adult} \mid \text{Stu}) \neq P(\neg \text{Adult} \vee \text{Stu})$ after Lewis (1976), who pointed out that there is no way of expressing a conditional probability by means of a standard connective of logic ; yet the statement "most students are adults" is better expressed by $P(\text{Adult} \mid \text{Stu}) \geq \alpha$ rather than by $P(\neg \text{Adult} \vee \text{Stu}) \geq \alpha$. For instance the contraposition law is not natural for conditional statements with exceptions (we cannot claim that most non-students are not adults, e.g. if the populations include many old people), nor is the transitivity acceptable, (e.g. most adults are employees, but almost no student is an employee) ; the lack of contraposition and transitivity laws is in accordance with conditional probabilities, not with probabilities of conditionals.

This view of probabilistic logic, which is in accordance with Kyburg (1974)'s claim that most of our knowledge consists in incomplete statistics, has not been very much developed on the proof-theoretic side ; however see Bacchus (1990). A straightforward method is to turn a set of weighted sentences (including conditional sentences) into a linear programming problem where variables are probabilities of interpretations (see e.g. Paass, 1988). Another, less obvious, path is to find inference rules for this type of logic. For instance the "quantified syllogism pattern" in which $P(r \mid p)$ is computed in terms of $P(p \mid q)$, $P(q \mid p)$ and $P(r \mid q)$ is an example of such an inference rule (see Dubois and Prade, 1988c). An extreme case of this approach is Adams (1975) conditional logics, recently revived by Pearl (1988). In this logic, conditional statements are of the form $P(q \mid p) \geq 1 - \epsilon$ where ϵ is infinitesimally small. A theory is then a pair (F, Δ) where F is a set of propositional sentences and Δ is a set of conditional

statements of the form $p \Rightarrow q$ which means $P(q | p) \geq 1 - \epsilon$. A plausible conclusion r from (F, Δ) is such that $P(r | F) \geq 1 - 0(\epsilon)$ for all P in the set of probability measures constrained by Δ . The following inference rules form the basis of a sound and complete deduction system :

$$\begin{array}{ll} \text{triangularity} & p \Rightarrow q, p \Rightarrow r \vdash p \wedge q \Rightarrow r \\ \text{"Bayes"} & p \Rightarrow q, p \wedge q \Rightarrow r \vdash p \Rightarrow r \\ \text{disjunction} & p \Rightarrow r, q \Rightarrow r \vdash p \vee q \Rightarrow r \end{array}$$

Interestingly, these properties are natural requirements for a well-behaved non-monotonic logic (Gabbay, 1985) ; non-monotonicity means that the arrival of new pieces of information may invalidate previously made conclusions. For instance 'flies' is a plausible conclusion of (F, Δ) with $F = \{\text{bird}, \neg \text{penguin} \vee \text{bird}\}$, $\Delta = \{\text{bird} \Rightarrow \text{flies}, \text{penguin} \Rightarrow \neg \text{flies}\}$ but is no longer a plausible conclusion of (F', Δ) where $F' = F \cup \{\text{penguin}\}$. It is well-known that in classical logic the consequence relation is reflexive ($p \Rightarrow p$), transitive ($p \Rightarrow q, q \Rightarrow r \vdash p \Rightarrow r$) and monotonic ($p \Rightarrow q \vdash p \wedge r \Rightarrow q$). The triangularity rule is a restricted form of monotonicity while the "Bayes"-like rule is a restricted form of transitivity. Other kinds of qualitative probabilistic logics exist, according to the way conditional statements are modelled ; $p \Rightarrow q$ may alternatively mean $P(q | p) > P(q)$ (Neufeld, 1990) or $P(q | p) > P(q | \neg p)$ (Wellman, 1990). Another point of view is to interpret $p \Rightarrow q$ as a conditional object $q | p$ whose probability is indeed a conditional probability. Schay (1968) has defined an algebra of such conditional events, where the truth-value of $q | p$ is defined as follow :

$$\begin{aligned} t(q | p) &= 1 \text{ if } t(p) = 1, t(q) = 1 \\ &= 0 \text{ if } t(p) = 1, t(q) = 0 \\ &= ? \text{ otherwise (undefined).} \end{aligned}$$

Very similar proposals have been independently reinvented by Goodman and Nguyen (1988) and Calabrese (1987). Given a Boolean algebra of propositions \mathfrak{B} , $q | p$ is viewed as the subset of \mathfrak{B} whose upper bound is $\neg p \vee q$ and lower bound is $p \wedge q$. For a systematic study of conditional objects based on the above three-valued semantics, see Dubois and Prade (1990a). Conditional objects also yield a non-monotonic logic whose properties are similar to Adams' logic (Dubois and Prade, 1991b).

III.2.2 - Possibilistic Logic

Instead of representing uncertain knowledge by means of a probability measure over interpretations, one may think of representing incomplete, vague knowledge over the same set of interpretations, defined from a given language. In other words we may start with a possibility

distribution over a set of interpretations which are atoms of a Boolean algebra of propositions, or over a set Ω of possible worlds. The degrees of possibility and necessity of any sentence is then defined by

$$\begin{aligned}\prod(p) &= \sup\{\pi(\omega) \mid \omega \models p\} \\ N(p) &= \inf\{1 - \pi(\omega) \mid \omega \models \neg p\}\end{aligned}$$

Conversely, one may think of starting with a given set of sentences, to which degrees of necessity or possibility are attached. Let us restrict to the case of weighted sentences of the form (p_i, α_i) with intended meaning $N(p_i) \geq \alpha_i$. Similarly to the case of probabilistic logics, constraints of the form $N(p_i) \geq \alpha_i$, $i = 1, n$ define a family of possibility measures π on Ω , such that

$$\inf\{1 - \pi(\omega) \mid \omega \models \neg p_i\} \geq \alpha_i, i = 1, n \quad (23)$$

There exists a unique possibility distribution π^* on Ω such that all π that satisfy (23) also satisfy $\pi \leq \pi^*$. π^* is the least specific possibility distribution obeying (23), i.e. where all the incompleteness of the available information is preserved. It is defined by

$$\pi^*(\omega) = \min_{i=1, n} \max(t_\omega(p_i), 1 - \alpha_i)$$

where $t_\omega(p_i)$ is the truth of p_i at world ω . As can be seen, in possibility theory the approach based on possible worlds and the approach based on sentences weighted by degrees of necessity are equivalent, through the principle of minimum specificity. A counterpart to the confirmation degree can be defined in possibilistic logic, namely to what extent a sentence p entails another formula q in the context of the knowledge base $\mathcal{S} = \{(p_i, \alpha_i) ; i = 1, n\}$. This is the conditional necessity degree defined by

$$N(q \mid p) = 1 - \prod(\neg q \mid p) \quad (24)$$

where $\prod(\neg q \mid p)$ is the greatest solution to the equation :

$$\prod(\neg q \wedge p) = \prod(\neg q \mid p) * \prod(p) \quad (25)$$

and N, \prod are calculated from the possibility distribution π^* (induced by \mathcal{S}). (25) is a particular case of the equation (10) defining the notion of conditionalization following Cox (1946). Operation $*$ can be chosen as $*$ = min or product (Dubois and Prade, 1990a). It is easy to verify that

$$N(q \mid p) > 0 \Leftrightarrow \prod(q \wedge p) > \prod(\neg q \wedge p)$$

It is interesting to define a notion of plausible entailment of q from p , denoted $p \Rightarrow q$, by requiring $N(q \mid p) > 0$. The following results have been established (Dubois and Prade, 1991c, b) :

- The entailment relation \Rightarrow is then a well-behaved non-monotonic relation (i.e. it satisfies reflexivity, triangularity, "Bayes", and disjunction). Actually \Rightarrow is equivalent to the notion of Shoham (1988)'s preferential entailment in a preference logic, i.e. a logic where the set of interpretations is equipped with a partial strict ordering relation $>$ that is transitive. This is the strict ordering induced by the possibility distribution π^* . The preferential entailment $p \models_{>} q$ means that q is true in all the maximal worlds where p is true, which is exactly expressed by the inequality $\prod(q \wedge p) > \prod(\neg q \wedge p)$.
- Adams' conditional logic can be related to possibility theory in a strong way, through the probabilistic interpretation of ordinal conditional functions which Spohn (1990) suggested (see section I.3 and II.1). Namely the following statements are equivalent :

- $P(q \mid p) \geq 1 - \varepsilon^n$, where ε is infinitesimal
- $\kappa(\neg q \mid p) = n$ for some κ -function
- $\prod(\neg q \mid p) = k^{-n} < 1$ for some $k > 1$
- $N(q \mid p) > 0$,

where $\prod(\neg q \mid p)$ is the conditional possibility in the sense of Dempster rule (see section II.1 formula (9)). It points out the fact that Adams' calculus of extreme probabilities borrows only little from the additive structure of probability, and that the rules of non-monotonic logics are more akin to possibility theory (i.e. the idea of ordering and preference) than to probability theory, contrary to what could be supposed at first glance.

From a proof theoretic point of view possibilistic logic is fairly advanced. A resolution rule has been proposed for necessity valued clauses (Dubois and Prade, 1987b). The corresponding refutation method has been proved sound and complete (Dubois et al., 1989). The mixed case, with both possibility and necessity valued clauses has been thoroughly studied (Lang et al., 1991).

III.3 - Fuzzy Logics

The name "fuzzy logics" has been used by several authors with different meanings : one rather often found understanding is a multiple-valued logic that accounts for the vagueness of predicates. This is not the way in which Zadeh (1979b) envisages fuzzy logics : according to

him fuzzy logics involves fuzzy truth-values computed on the basis of possibility distributions that model the meaning of natural language sentences ; a more restricted meaning of fuzzy logics refers to fuzzy rule-based controllers (Mamdani, 1977), currently applied to the automatic control of industrial processes. Lastly Ruspini (1991) has recently put forward still another view of fuzzy logic, relating it to the problem of indiscernibility at least as much as to the problem of vagueness.

The first proposal of a multiple-valued logic approach of vagueness is by Goguen (1969), who is motivated by the sorites paradox of classical logic, namely examples like the following statements

- A) "A person of one year-old is young"
- B) "A person of n -year-old is young" implies "A person of $(n + 1)$ -year-old is young"
- C) "A person of 100-years-old is not young"

which become contradictory if expressed in classical logic, while they agree with commonsense. One way out of this paradox is to admit that young is a vague predicate and that young(n) (that stands for "A person of n -year-old is young") can take intermediary truth-values between True and False, usually modelled by means of numbers between 1 and 0. Then it is possible to express that young($n + 1$) is slightly less true than young(n) so that young(100) can become false without contradicting B and A.

Using multiple-valued logics to handle vagueness leads to several problems. A first problem is how to define a well-formed formula that can express "John is young is rather true", where "rather true" is some intermediary truth-value like 0.8. Goguen (1969) did not consider this problem because he kind of gave up the idea that a syntax for a logic of vagueness might make sense; for him vagueness and intermediary truth-values were basically semantic notions. A way out of this problem is to introduce intermediary truth-values inside the language. This is what has been done by Pavelka (1979) and Novak (1990), who prove completeness results for special structures of the truth-value set.

However using a multiple-valued logic with formulas weighted by truth-values does not sound very natural because as pointed out earlier, if knowing the size of John, say 1.70 m, it makes sense to evaluate the truth of the statement "tall(John)" by means of $\mu_{\text{tall}}(1.70)$, it no longer makes sense to guess John's size (precisely) from the fact that tall(John) is true with degree 0.8. A first reason is that nobody can measure truth on the unit interval and express one's knowledge under this form. A second reason is that if it would be possible to produce truth-values as part of our knowledge, the restriction to precise truth-value implicitly assumes that

everything about the world is known, so that truth-values can be computed. Usually, the truth of vague statements is partially unknown, because our knowledge is incomplete.

The proposal that formulas weighted by truth-values are the regular extension of the classical case where formulas taken as axioms are weighted by truth-value 1 can be questioned. Indeed it underlies the assumption that "tall(John) is true" means $\mu_{\text{tall}}(\text{size}(\text{John})) = 1$. Assume 'tall' is such that $\exists s \in S$ (where S denotes the set of sizes), $\forall u \geq s$, $\mu_{\text{tall}}(u) = 1$, as usually assumed in fuzzy set theory, then "tall(John) is true" is equivalent to a classical statement $P(\text{John})$ where P is the non-fuzzy predicate whose extension is the set of sizes greater than or equal to s . And indeed, the statement "tall(John) is true" may sometimes, in natural language, express the idea that "John is a prototype of a tall man". However, this approach leads to interpret all true fuzzy statements as non-fuzzy ones. Another possible interpretation is the following :

"tall(John) is true" is equivalent to "what is known is just : 'tall(John)'"

This means that all sizes that are compatible whatsoever with 'tall' are more or less possible sizes for John. Particularly, it only implies that "John is not a prototype of 'not tall'". This is Zadeh's view of fuzzy logic, whereby all statements are interpreted as possibility distributions (soft constraints) on the range of parameters of interest. Here the membership function μ_{tall} is viewed as a possibility distribution on the size of John. Again, this view is basically semantic and tells us nothing about how to specify the syntax of a language for reasoning with vague predicates, i.e. how to devise a "logic of vagueness".

Zadeh (1979b) views truth-values as modifiers of fuzzy predicates, i.e. another way of expressing linguistic hedges such as "very", "rather" etc.... Given the membership function of 'tall', the membership function of 'very tall' is assumed to be modelled by $f(\mu_{\text{tall}})$ where f is a modifier, i.e. a function from $[0,1]$ to $[0,1]$ that modifies the shape of μ_{tall} . Then Zadeh (1979b) views natural language statements like "John is tall is very true" as meaning "John is very tall". So, the modifier f can be viewed as a (fuzzy) truth-value, i.e. a distribution of possible truth-values on the unit interval. And the identity between "tall(John) is true" and "John is tall" leads to interpret the truth-value "true" as the identity function on $[0,1]$, i.e. $f(u) = u$. This view of a logic of vagueness is quite different from the idea of a multiple-valued logic where numbers between 0 and 1 may directly model intermediary truth-values. Anyway note that Lukasiewicz logic axioms (Wajsberg, 1935) do not involve syntactical counterparts of truth-values, but are expressed in a language similar to propositional logic, even if it is an infinite-valued logic at the semantic level.

This distinction between the two understandings of "p is true" where p is a vague proposition may help understand objections of scholars like Kit Fine (1975) against many-valued logics as capturing vagueness. For Fine, "A vague sentence is true if it is true for all admissible and complete specifications" of the sentence ; in other words a vague sentence is true if any non-vague representation of the vague sentence is true. Such a representation is obtained by assigning only definite truth-values "true" or "false". But if we adopt the idea that the truth of "tall(John)" means that John is a prototype of "tall man", accepting this state of fact does not depend on the choice of the threshold between "tall" and "not tall". And even if "tall" is vague the sentence 'tall(John) \vee \neg tall(John)' will be true, and 'tall(John) \wedge \neg tall(John)', whether this threshold is chosen or not, because using the words of Fine : "the boundary of the one shifts with the boundary of the other". Zadeh's view of vagueness never presupposes that there is a moving threshold between the extension of a vague predicate and the one of its negation. "tall(John) is true" is interpreted as a flexible constraint, not to an implicit decision about the location of John's size with respect to a moving threshold.

Note that Fine (1975) uses the terminology "indefinite truth" to denote a third truth-value different from "true" and "false". This terminology is ambiguous since it is not clear whether "indefinite" means "unknown" or "intermediary". Especially if a proposition p obeys the rules of classical logic the truth of p can be unknown, as well as the truth of \neg p, but it cannot be intermediary. Moreover $p \wedge \neg p$ is false and $p \vee \neg p$ is true, even if the truth of p and $\neg p$ is unknown, while $p \wedge p$ has unknown truth. As Fine (1975) we conclude again that the truth-function is not truth-functional with respect to the three truth-values 'true', 'false', and 'undefined'. But unknown truth is completely captured by possibilistic logic (Section III.2.2), without requiring truth-functionality. On the contrary intermediate truth makes no sense for non-vague predicates, and using Zadeh's understanding of fuzzy statements, 'tall(John) \wedge \neg tall(John)' can be considered as semantically equivalent to 'medium-size(John)' ; it implies that John is neither a prototype of "tall" nor of its contrary, noticing that "not to be a prototype of p" is not the same as "to be a prototype of \neg p".

As said earlier, the problem of devising a genuine logic of vague predicates without resorting to reified truth-values on the unit interval is still open. At this point the current activity rather aims at getting computational models of vague complex statements of natural language ; this is the purpose of Zadeh (1978b)'s translation tool PRUF which systematically turns natural language sentences into possibility distributions acting as flexible constraints. The study of various interpretations of fuzzy "if... then" statements and the corresponding multiple-valued implications has been considered for a long time (see Dubois and Prade, 1991d).

Ruspini (1991)'s notion of fuzzy logic, is based on the notion of possible world, and on

a metric structure that quantifies similarity between possible worlds. For Ruspini, given two classical propositions p and q , the extent to which p implies q is the minimal amount of "stretching" that is required to include the set of worlds in which p is true in the neighborhood of the worlds where q is true. More formally if Ω is the set of possible worlds, if $\omega \models p$ means p is true in $\omega \in \Omega$, then the following metric counterpart of Carnap's confirmation degree is

$$I(q \mid p) = \inf_{\omega \models p} \sup_{\omega' \models q} \mu_R(\omega, \omega')$$

where R is a fuzzy proximity relation, that expresses indiscernibility on Ω . In order to clarify the meaning of $I(q \mid p)$, let us consider the case when R is a standard equivalence relation. Let q^* be the upper approximation of q in the sense of R (see Section I.5), i.e. $\omega \models q^* \Leftrightarrow \exists \omega', \omega R \omega'$ and $\omega \models q$; then $I(q \mid p) = 1$ corresponds to the semantic entailment $p \models q^*$. We get very close to the logic of rough sets developed by Fariñas del Cerro and Orłowska (1985). Clearly, if R becomes a similarity relation (for instance $1 - \mu_R(\cdot, \cdot)$ corresponds to a distance on Ω), then q^* is a fuzzy predicate with membership function $\sup_{\omega' \models q} \mu_R(\cdot, \omega')$. Then the extent to which $\omega \models q^*$ is true becomes a matter of degree even if q is not vague. It reflects the distance between ω and the worlds in which q is true. We are not far from Weston (1987)'s approximate truth, but no longer very close to Zadeh's notion of fuzzy logic. Especially, defining degrees of truth $t^\omega(q) = \sup_{\omega' \models q} \mu_R(\omega, \omega')$, we no longer have truth functionality, e.g. $t^\omega(q \wedge q') \leq \min(t^\omega(q), t^\omega(q'))$, even if R is not fuzzy, i.e. even if $t^\omega(\cdot)$ is 2-valued. Indeed, in rough set theory we only have the inclusion $(q \wedge q')^* \subseteq q^* \wedge q'^*$, so that we may have $t^\omega(q) = 1$, $t^\omega(q') = 1$ but $t^\omega(q \wedge q') = 0$.

IV - Conclusion

There are many facets to the imperfection of information, and it is claimed here that the traditional framework of probability theory needs to be enlarged (rather than disputed) in order to account for these facets. Among interesting aspects, incompleteness, indiscernibility and vagueness may disturb the ideal picture of the standard probabilistic setting. As a result, notions such as conditioning and updating no longer coincide, and the problem of combining imperfect evidence becomes central. On the side of logic, logics of uncertainty in which the truth of sentences cannot be established due to incomplete or contradictory evidence, have been opposed to logics of vagueness where truth-levels can be intermediary as reflecting imperfect compatibility between sentences and complete evidence. Logics of uncertainty apply to classical sentences and degrees of uncertainty are not compositional; they either rely on an additive structure as in the case of incomplete statistical evidence, or on an ordinal structure as in the case of possibilistic logic where degrees of certainty express a reluctance to delete sentences in the

presence of contradictions (Dubois and Prade, 1991e). As for logics of vagueness, it has been suggested that a direct handling of numerical truth-values in the syntax may be counter-intuitive and anyway does not correspond to Zadeh's original view of reasoning with vague statements. In some sense we are still with Goguen (1969)'s doubts regarding a full-flavored logic of vagueness, although the problem of meaning representation of vague knowledge, and automated procedures for inference from these representations has drastically progressed.

References

- Adams E.W. (1975) *The Logic of Conditionals*. Reidel, Dordrecht.
- Adams E.W., Levine H.P. (1975) On the uncertainties transmitted from premises to conclusions in deductive inferences. *Synthese*, 30, 429-460.
- Bacchus F. (1990) *Representing and Reasoning with Probabilistic Knowledge – A Logical Approach to Probabilities*. MIT Press, Cambridge, MA & London.
- Bezdek J.C. (1981) *Pattern Classification with Fuzzy Objective Function Models*. Plenum Press, New York.
- Boole G. (1854) *An Investigation of the Laws of Thought on which are Founded the Mathematical Theory of Logic and Probabilities*. MacMillan. (Reprinted by Dover, New York, 1958).
- Buchanan B.G., Shortliffe E.H. (1984) *Rule-Based Expert Systems – The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison-Wesley, Reading.
- Calabrese P. (1987) An algebraic synthesis of the foundations of logic and probability. *Information Sciences*, 42, 187-237.
- Carnap R. (1950) *Logical Foundations of Probability*. Routledge & Kegan Paul, London.
- Chateauneuf A. (1988a) Uncertainty aversion and risk aversion in models with nonadditive probabilities. In : *Risk, Decision and Rationality* (B.R. Munier, ed.), Reidel, Dordrecht, 615–629.
- Chateauneuf A. (1988b) Decomposable measures, distorted probabilities and concave capabilities. *FUR-IV Conf. "Foundations of Utility and Risk Theories*, Budapest, April.
- Choquet G. (1953) Theory of capacities. *Ann. Inst. Fourier (Grenoble)*, 5(4), 131–295.
- Cox R.T. (1946) Probability, frequency and reasonable expectation. *American Journal of Physics*, 14, 1–13.
- De Campos L.M., Lamata M.T., Moral S. (1990) The concept of conditional fuzzy measure. *Int. J. of Intelligent Systems*, 5, 237-246.
- De Finetti B. (1937) *La prévision : ses lois logiques, ses sources subjectives*. *Ann. Inst. Poincaré* 7, 1937, 1-68. Translated in : *Studies in Subjective Probability* (J. Kyburg, H.E. Smokler, eds.), Wiley, New York, 1964.

- Dempster A.P. (1967) Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.*, 38, 325–339.
- Dubois D. (1986a) Belief structures, possibility theory and decomposable confidence measures on finite sets. *Computers and Artificial Intelligence (Bratislava)*, 5, 403–416.
- Dubois D. (1986b) Generalized probabilistic independence and its implications for utility. *Operations Research Letters*, 5, 255–260.
- Dubois D., Lang J., Prade H. (1989) Automated reasoning using possibilistic logic : semantics, belief revision and variable certainty weights. *Proc. of the 5th Workshop on Uncertainty in Artificial Intelligence*, Windsor, Ontario, August 18-20, 81-87.
- Dubois D., Prade H. (1980) *Fuzzy Sets and Systems : Theory and Applications*. Mathematics in Sciences and Engineering Series, Vol. 144, Academic Press, New York.
- Dubois D., Prade H. (1982) A class of fuzzy measures based on triangular norms. *Int. J. of General Systems*, 8, 43–61.
- Dubois D., Prade H. (1985a) Evidence measures based on fuzzy information. *Automatica*, 21, 547–562.
- Dubois D., Prade H. (1985b) A review of fuzzy set aggregation connectives. *Information Sciences*, 36, 85–121.
- Dubois D., Prade H. (1985c) (with the collaboration of Farreny H., Martin-Clouaire R., Testemale C.) *Théorie des Possibilités – Applications à la Représentation des Connaissances en Informatique*. Masson, Paris. 2nd revised and augmented edition, 1987). English translation : *Possibility Theory – An Approach to the Computerized Processing of Uncertainty*, Plenum Press, New York, 1988.
- Dubois D., Prade H. (1986a) The principle of minimum specificity as a basis for evidential reasoning. In : *Uncertainty in Knowledge-Based Systems (Inter. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Paris, France, June 30-July 4, 1986)* (B. Bouchon, R.R. Yager, eds.), *Lecture Notes in Computer Science*, Springer Verlag, Berlin, 75–84.
- Dubois D., Prade H. (1986b) A set-theoretic view of belief functions : logical operations and approximations by fuzzy sets. *Int. J. of General Systems*, 12, 193–226.
- Dubois D., Prade H. (1987a) The mean value of a fuzzy number. *Fuzzy Sets and Systems*, 24, 279–300.
- Dubois D., Prade H. (1987b) Necessity measures and the resolution principle. *IEEE Trans. on Systems, Man and Cybernetics*, 17, 474-478.
- Dubois D., Prade H. (1988a) Modelling uncertainty and inductive inference : a survey of recent non-additive probability systems. *Acta Psychologica*, 68, 53–78.
- Dubois D., Prade H. (1988b) An introduction to possibilistic and fuzzy logics. In : *Non-Standard Logics for Automated Reasoning* (P. Smets, A. Mamdani, D. Dubois, H. Prade, eds.), Academic Press, New York, 287–326.

- Dubois D., Prade H. (1988c) On fuzzy syllogisms. *Computational Intelligence (Canada)*, 4(2), 171–179.
- Dubois D., Prade H. (1988d) Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence (Canada)*, 4(4), 244-264.
- Dubois D., Prade H. (1989) Fuzzy sets, probability and measurement. *Europ. J. of Operational Research*, 40, 135–154.
- Dubois D., Prade H. (1990a) The logical view of conditioning and its applications to possibility and evidence theories. *Int. J. of Approximate Reasoning*, 4, 23-46.
- Dubois D., Prade H. (1990b) Rough fuzzy sets and fuzzy rough sets. *Int. J. of General Systems*, 17, 191-209.
- Dubois D., Prade H. (1991a) Updating with belief functions, ordinal conditional functions and possibility measures. In : *Uncertainty in Artificial Intelligence Vol. 6* (P.P. Bonissone, M. Henrion, L.N. Kanal, J.F. Lemmer, eds.), North-Holland, Amsterdam, to appear.
- Dubois D., Prade H. (1991b) Conditional objects and non-monotonic reasoning. *Proc. of the 2nd Inter. Conf. on Principles of Knowledge Representation and Reasoning (KR'91)*, Cambridge, MA, April 22-25, 1991 (J. Allen, R. Fikes, E. Sandewall, eds.), Morgan Kaufmann, to appear. *12th Inter. Joint Conf. on Artificial Intelligence (IJCAI-91)*, Sydney, Aus.
- Dubois D., Prade H. (1991c) Possibilistic logic, preference models, non-monotonicity and related issues. In : *Tech. Report IRIT/90-54/R*, IRIT, Univ. P. Sabatier, Toulouse, France, December 1990. *12th Inter. Joint Conf. on Artificial Intelligence(IJCAI-91)*, Sydney, Aus.
- Dubois D., Prade H. (1991d) Fuzzy sets in approximate reasoning – Part 1 : Inference with possibility distributions. *Fuzzy Sets and Systems, 25th Anniversary Memorial Volume*, to appear.
- Dubois D., Prade H. (1991e) Epistemic entrenchment and possibilistic logic. In : *Tech. Report IRIT/90-2/R*, IRIT, Univ. P. Sabatier, Toulouse, France. To appear in *Artificial Intelligence*.
- Dubucs J.P. (1989) Logiques non-classiques. *Encyclopedia Universalis*, 977-992.
- Fagin R., Halpern J.Y. (1989a) Uncertainty, belief and probability. *Proc. of the 11th Inter. Joint Conf. on Artificial Intelligence (IJCAI-89)*, Detroit, Michigan, 1161-1167
- Fagin R., Halpern J.Y. (1989b) A new approach to updating beliefs. *Research Report RJ 7222*, IBM, Research Division, San Jose, Ca..
- Fariñas del Cerro L., Orłowska E. (1985) DAL – A logic for data analysis. *Theoretical Computer Science*, 36, 251-264.
- Fariñas del Cerro L., Prade H. (1986) Rough sets, twofold fuzzy sets and modal logic – Fuzziness in indiscernibility and partial information. In : *The Mathematics of Fuzzy Systems* (A. Di Nola, A.G.S. Ventre, eds.), Verlag TÜV Rheinland, Köln, 103–120.
- Fenstad J.E. (1967) Representations of probabilities defined on first order languages.

- In : Sets, Models and Recursion Theory (J.N. Crossley, ed.), North-Holland, Amsterdam, 156–172.
- Fine K. (1975) Vagueness, truth and logic. *Synthese*, 30, 265-300.
- Fine T.L. (1973) *Theories of Probability*. Academic Press, New York.
- Fishburn P.C. (1986a) The axioms of subjective probability. *Statistical Science*, 1, 335–358.
- Fishburn P.C. (1986b) Interval models for comparative probability on finite sets. *J. of Mathematical Psychology*, 30, 221–242.
- Gabbay D.M. (1985) Theoretical foundations for non-monotonic reasoning in expert systems. In : *Logics and Models of Concurrent Systems* (K.R. Apt, ed.), Springer Verlag, Berlin, 439-457.
- Gärdenfors P. (1988) *Knowledge in Flux – Modeling the Dynamics of Epistemic States*. The MIT Press, Cambridge, MA & London.
- Gärdenfors P., Hansson B., Sahlin N.E. (Eds.) (1983) *Evidentiary Value : Philosophical, Judicial and Psychological Aspects of a Theory*. CWK Gleerups, Lund, Library of Theoria n° 15.
- Gil M.A. (1988) Probabilistic-possibilistic approach to some statistical problems with fuzzy experimental observations. In : *Combining Fuzzy Imprecision with Probabilistic Uncertainty in Decision-Making* (J. Kacprzyk, M. Fedrizzi, eds.), *Lecture Notes in Economics and Mathematical Systems*, Vol. 310, Springer Verlag, Berlin, 286-306.
- Gilboa I. (1987) Expected utility with purely subjective non-additive probabilities. *J. Math. Econom.*, 16, 65–88.
- Giles R. (1982) Foundations for a theory of possibility. In : *Fuzzy Information and Decision Processes* (M.M. Gupta, E. Sanchez, eds.), North-Holland, Amsterdam, 183–195.
- Goguen J.A. (1969) The logic of inexact concepts. *Synthese*, 19, 1-36.
- Goodman I.R., Nguyen H.T. (1988) Conditional objects and the modeling of uncertainties. In : *Fuzzy Computing – Theory, Hardware, and Applications* (M.M. Gupta, T. Yamakawa, eds.), North-Holland, Amsterdam, 119-138.
- Halmos P. (1950) *Measure Theory*. Van Nostrand.
- Hisdal E. (1988) Are grades of membership probabilities ?. *Fuzzy Sets and Systems*, 25, 325–348.
- Höhle U. (1988) Quotients with respect to similarity relations. *Fuzzy Sets and Systems*, 27, 31-44.
- Hughes G.E., Cresswell M.J. (1968) *An Introduction to Modal Logic*. Methuen, London.
- Jaffray J.Y. (1990) Bayesian updating and belief functions. *Proc. of the 3rd Inter. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'90)*, Paris, July, 449-451 (published by ENSTA, Paris).
- Jardine N., Sibson R. (1977) *Mathematical Taxonomy*. Wiley, New York.
- Jeffrey R. (1983) *The Logic of Decision* (2nd ed.). University of Chicago Press, Chicago, IL.

- Klement E.P. (1982) Construction of fuzzy σ -algebras using triangular norms. *J. of Mathematical Analysis and Applications*, 85, 543–566.
- Klir G.J. (ed.) (1987) Special Issue on "Measures of Uncertainty". *Fuzzy Sets and Systems*, 24(2).
- Kruse R., Meyer K.D. (1987) *Statistics with Vague Data*. D. Reidel, Dordrecht.
- Kyburg H.E., Jr. (1974) *The Logical Foundations of Statistical Inference*. D. Reidel, Dordrecht.
- Kyburg H.E., Jr. (1987) Bayesian and non-Bayesian evidential updating. *Artificial Intelligence*, 31, 271–293.
- Lang J., Dubois D., Prade H. (1991) A logic of graded possibility and certainty coping with partial inconsistency. In : Tech. Report IRIT/90-54/R, IRIT, Univ. P. Sabatier, Toulouse, France, 1990. Submitted.
- Lehrer K., Wagner C. (1981) *Rational Consensus in Science and Society*. D. Reidel Publishing Company, Boston.
- Lewis D. (1976) Probabilities of conditionals and conditional probabilities. *Phil. Rev.*, 85, 297–315.
- Løf J. (1963) Semantic representations of the probability of formulas in formalized theories. *Studia Logica*, 14, 183–194.
- Mamdani E.H. (1977) Application of fuzzy logic to approximate reasoning using linguistic systems. *IEEE Trans. on Computers*, 26, 1182–1191.
- Menger K. (1951) Probabilistic theories of relations. *Proc. Nat. Acad. Sci. USA*, 37, 178–180.
- Murofushi T., Sugeno M. (1989) An interpretation of fuzzy measures and the Choquet integral as an integral with respect to a fuzzy measure. *Fuzzy Sets and Systems*, 29, 201–227.
- Neufeld E. (1990) A probabilistic commonsense reasoner. *Int. J. of Intelligent Systems*, 5, 565–594.
- Nguyen H.T. (1978) On random sets and belief functions. *J. of Mathematical Analysis and Applications*, 65, 531–542.
- Nilsson N. (1986) Probabilistic logic. *Artificial Intelligence*, 28, 71–87.
- Norwich A.M., Turksen I.B. (1982) The fundamental measurement of fuzziness. In : *Fuzzy Set and Possibility Theory : Recent Development* (R.R. Yager, ed.), Pergamon Press, Oxford, 49–50.
- Novak V. (1990) On the syntactico-semantic completeness of first-order fuzzy logic – Part I : Syntax and semantics. *Kybernetika*, 26(1), 47–66 ; Part II : Main results, *Kybernetika*, 26(2), 134–154.
- Paass G. (1988) Probabilistic logic. In : *Non-Standard Logics for Automated Reasoning* (P. Smets, E.H. Mamdani, D. Dubois, H. Prade, eds.), Academic Press, New York, 213–251.
- Pavelka J. (1979) On fuzzy logic. *Zeitschr. f. Math. Logik und Grundlagen d. Math.*, 25,

- Part : 45–72 ; Part II : 119–134 ; Part III : 447–464.
- Pawlak Z. (1982) Rough sets. *Int. J. of Computer and Informatics Sciences*, 11, 341–356.
- Pearl J. (1988) *Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference*. Morgan Kaufman, San Mateo, CA.
- Poincaré H. (1902) *La Science et l'Hypothèse*. Flammarion, Paris.
- Prade H. (1985) A computational approach to approximate reasoning with applications to expert systems. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 7, 260–283. Corrections, 7, 747–748.
- Reichenbach H. (1949) *The Theory of Probability*. University of California Press, Berkeley and Los Angeles.
- Ruspini E.H. (1991) On the semantics of fuzzy logic. *Int. J. of Approximate Reasoning*, 5, 45-88.
- Schay G. (1968) An algebra of conditional events. *J. Math. Anal. & Appl.*, 24, 334-344.
- Schmeidler D. (1986) Integral representation without additivity. *Proc. Amer. Math. Soc.*, 97, 255–261.
- Schweizer B., Sklar A. (1983) *Probabilistic Metric Spaces*. North-Holland, Amsterdam.
- Shackle G.L.S. (1961) *Decision, Order and Time in Human Affairs*. (2nd edition) Cambridge University Press, Cambridge.
- Shafer G. (1976) *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, N.J..
- Shafer G. (1978) Non-additive probabilities in the work of Bernoulli and Lambert. *Archive for History of Exact Sciences*, 19, 309-370.
- Shafer G. (1986) The combination of evidence. *Int. J. of Intelligent Systems*, 1, 155-180.
- Shafer G. (1990) The unicity of probability. In : *Acting Under Uncertainty : Multidisciplinary Conceptions* (G.M. von Furstenberg, ed.), Kluwer Academic Pub., Boston.
- Shoham Y. (1988) *Reasoning About Change – Time and Causation from the Standpoint of Artificial Intelligence*. The MIT Press, Cambridge, Mass.
- Smets P. (1978) Un modèle mathématico-statistique simulant le processus du diagnostic médical. Doctoral Dissertation, Free University of Brussels, Presses Universitaires de Bruxelles. Available through University Microfilm International, 30-32 Mortimer Street, London W1N 7RA, Thesis 80-70,003).
- Smets P. (1981) The degree of belief in a fuzzy event. *Information Sciences*, 25, 1–19.
- Smets P. (1988) Belief functions. In : *Non-Standard Logics for Approximate Reasoning* (P. Smets, A. Mamdani, D. Dubois, H. Prade, eds.), Academic Press, New York, 253-286.
- Smith C.A.B. (1961) Consistency in statistical inference and decision. *J. Royal Statist. Soc.*, B-23, 1–37.
- Spohn W. (1988) Ordinal conditional functions : a dynamic theory of epistemic states.

- In : Causation in Decision, Belief Change, and Statistics (W.L. Harper, B. Skyrms, eds.), Kluwer Academic Publ., 105-134.
- Spohn W. (1990) A general non-probabilistic theory of inductive reasoning. In : Uncertainty in Artificial Intelligence 4 (R.D. Shachter, T.S. Levitt, L.N. Kanal, J.F. Lemmer, eds.), North-Holland, Amsterdam, 149-158.
- Sugeno M. (1977) Fuzzy measures and fuzzy integral : a survey. In : Fuzzy Automata and Decision Processes (M.M. Gupta, G.N., Saridis, B.R. Gaines, eds.), North-Holland, Amsterdam, 89-102.
- Suppes P. (1974) The measurement of belief. *J. of Royal Statist. Soc.*, B-26, 160-191.
- Suppes P., Zanotti M. (1977) On using random relations to generate upper and lower probabilities. *Synthese*, 36, 427-440.
- Turksen I.B. (1991) Measurement of membership functions and their acquisition. *Fuzzy Sets and Systems*, Silver Anniversary Issue, to appear.
- Valverde L. (1985) On the structure of F-indistinguishability operators. *Fuzzy Sets and Systems*, 17, 313-328.
- Wagner C.G. (1989) Consensus for belief functions and related uncertainty measures. *Theory and Decision*, 26, 295-304.
- Wajsberg M. (1935) Beiträge zum Metaaussagenkalkül I. *Monatshefte für Mathematik und Physik*, 42, 221-242.
- Wakker P.P. (1989) Continuous subjective expected utility with nonadditive probabilities. *J. Math. Econom.*, 18, 1-27.
- Weber S. (1984) \perp -Decomposable measures and integrals for Archimedean t-conorms \perp . *J. of Mathematical Analysis and Applications*, 101, 114-138.
- Weber S. (1988) Conditional measures based on Archimedean semigroups. *Fuzzy Sets and Systems*, 27, 63-72.
- Wellman M.P. (1990) Fundamental concepts of qualitative probabilistic networks. *Artificial Intelligence*, 44, 257-303.
- Weston T. (1987) Approximate truth. *J. Philos. Logic*, 16, 203-227.
- Williams P. (1980) Bayesian conditionalization and the principle of minimum information. *British J. for the Philosophy of Science*, 31, 131-144.
- Wong S.K.M. (1991) Propagation of preference relations in qualitative inference networks. 12th Inter. Joint Conf. on Artificial Intelligence, Sydney, Aus.
- Yaari M.E. (1987) The dual theory of choice under risk. *Econometrica*, 55, 95-115.
- Yager R.R. (1984) Probabilities from fuzzy observations. *Information Sciences*, 32, 1-31.
- Zadeh L.A. (1965) Fuzzy sets. *Information and Control*, 8, 338-353.
- Zadeh L.A. (1968) Probability measures of fuzzy events. *J. of Mathematical Analysis and Applications*, 23, 421-427.
- Zadeh L.A. (1971) Similarity relations and fuzzy orderings. *Information Sciences*, 3, 177-200.

- Zadeh L.A. (1975) The concept of a linguistic variable and its application to approximate reasoning. *Information Sciences*, Part I : 8, 199-249 ; Part II : 8, 301-357 ; Part III : 9, 43-80.
- Zadeh L.A. (1978a) Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1, 3–28.
- Zadeh L.A. (1978b) PRUF : a meaning representation language for natural languages. *Int. J. of Man-Machine Studies*, 10(4), 395-460.
- Zadeh L.A. (1979a) Fuzzy sets and information granularity. In : *Advances in Fuzzy Set Theory and Applications* (M.M. Gupta, R.K. Ragade, R.R. Yager, eds.), North-Holland, Amsterdam, 3-18.
- Zadeh L.A. (1979b) A theory of approximate reasoning. In : *Machine Intelligence*, Vol. 9 (J.E. Hayes, D. Michie, L.I. Mikulich, eds.), Elsevier, New York, 149–194.
- Zadeh L.A. (1982) Test-score semantics for natural languages and meaning representation via PRUF. In : *Empirical Semantics* (B.B. Rieger, ed.), Brockmeyer, Bochum, Germany, 281-349.
- Zadeh L.A. (1987) A computational theory of dispositions. *Int. J. of Intelligent Systems*, 2, 39-63.