Numerical Representation of Uncertainty.

Philippe Smets[1]
IRIDIA
Université Libre de Bruxelles

1. Imperfect Data.

a) Three types. Information is perfect when it is consistent, precise and certain. Imperfection can thus be due to inconsistency, imprecision or uncertainty. When perfect information describes a world, it should uniquely determine it.

When no world fits with the information, we speak of **inconsistency**. Ordering or weighting inconsistency is usually not considered, except when authors try to restore consistency and can chose among several alternatives to rebuild it. In that case the selection of the appropriate solution can be achieved by using some ordering among the worlds, some being 'less inconsistent' with the information than other (Gärdenfors, 1988, Gabbay and Hunter, 1991). Usually the ordering reflects the 'amount of modification' that must be applied to the information so that consistency can be restored.

When more than one world fit with the information, we speak of **imprecision**. In its crudest form, an imprecise information is categorical and will only allow its user to know that the actual world belongs to a given set D of worlds. More elaborate forms of imprecise information are obtained with the introduction of some ordering among the worlds in D. The ordering reflects the similarity between the imprecise information and the information that characterizes each world in D. Classically such an ordering results from the fuzzy nature of some of the components of the imprecise information. If I know that Peter has more than 1 child, D will contain the worlds where Peter has 2, 3, 4... children. None of the worlds in D will be more similar to the imprecise information 'more than 1 child' than any other. Instead if I know that Peter has a few children, worlds where Peter has 2, 3, 4 children are more similar to the imprecise information 'a few children' than worlds where Peter has 7, 10, 13... children.

Beside imprecision, another ordering might be described, that we call **uncertainty**. It appears when another ordering is introduced atop of the imprecision. The uncertainty weights usually express the opinion of an agent about which of the possible worlds in D seems better 'supported' for being the actual world. Usually, the imprecise information is

[1] Research work has been partly supported by the Action de Recherches Concertées BELON funded by a grant from the Communauté Française de Belgique and the ESPRIT III, Basic Research Action 6156 (DRUMS II) funded by a grant from the Commission of the European Communities.

taken in its categorical form, hence the set D is just a set, without superposed similarity related ordering. The uncertainty related ordering on the worlds of D is usually achieved by a measure of uncertainty, like the probability measure. The set D can be seen then as the domain of the probability measure. Generalizations of uncertainty measures to sets D endowed with a similarity ordering have been developed in order to describe concepts like ‘the probability or the belief that a fuzzy event occurs’, but they will not be studied here (Zadeh, 1968, Smets 1981, 1982a)

b) Interrelation. Inconsistency and imprecision are properties related to the content of the statement that bears the information. It express the similarity or the compatibility of the information with the possible worlds. Uncertainty is a meta-property that express the opinion of the user about which world is the actual world. Uncertainty is a property of the relation between the information and a knowledge about the worlds, based on an integration of the available information.

Imprecision results in the construction of a domain D in which the actual world is known to belong, whereas uncertainty expresses the user’s opinions about which of these worlds is indeed the actual world. In Peter’s children example, I build D based on the imprecise information ‘more than one child’, but I might nevertheless give more belief to the fact that Peter has three children than two, etc...

Imprecise information tells about which worlds ***might*** be the actual world, uncertain information tells about which world among those that might be the actual world ***is*** indeed the actual world. In a very crude way, imprecision tells about the domain on which uncertainty measure is built.

Imprecision always refers to incompleteness of information or lack of specificity, and is expressed, in its crudest form, by disjunctions in logic or subsets of mutually exclusive states in set theory. When such sets expressing incomplete information are ordered by weighting the possible states, one gets fuzzy sets, possibility distributions and the like. Uncertainty refers to a state of partial belief modeled by attaching to each proposition (or event or set of possible states) a weight of support expressing Your confidence in the truth of that proposition (the occurrence of that event, the claim that the qualified set of possible states contains the real state of affairs). Uncertainty may be due to imprecision, but not necessarily so: inconsistency, randomness and the like are other sources of uncertainty. Noticeably a probabilistic representation of uncertainty generally cannot express imprecision understood as lacking, incomplete information.

c) Illustrating the difference. Going back to the case where Peter has 'a few children', a world where Peter has three children is more ‘possible’ (more similar to the imprecise information) then a world where Peter has five children. Nevertheless I might believe more in the fact that Peter has five children than three. Nothing forbid such an opinion. This discrepancy just enhances the difference between the concepts of imprecision and uncertainty.

To further illustrate this difference, consider the following two situations :

1. John has at least two children and I am sure about it.
2. John has three children but I am not sure about it.

In case 1, the number of children is imprecise but certain. In case 2, the number of children is precise but uncertain. Both aspects coexist but are distinct. Often the more imprecise you are, the most certain you are, and the more precise, the less certain. There seems to be some **Information Maximality Principle** that requires that some kind of 'product' of precision and certainty cannot be beyond a certain critical level. In its neighborhood, any increase in one is balanced by a decrease in the other (Zadeh, 1973).

d) Paper's content. In this paper, we focus essentially on the quantified representations of uncertainty, considering weighted imprecision only in its relation to uncertainty. We do not cover the symbolic approaches. The survey is not exhaustive. We restrict ourselves on the theories with well structured foundations, leaving aside ad hoc theories like those based on the 'certainty factors'...

We will successively define the concept of support, the frame on which uncertainty is defined and the concepts of actual worlds, of 'evidential corpus' and of the 'support state' that describes the agent's opinion about which of the possible worlds is the actual world (section 2). We present the distinction between generic and factual information (section 3). We then proceed by describing various theories for representing quantified uncertainty (sections 4 to 8).

2. The concept of support.

Choosing an appropriate word to cover what quantified uncertainty aims at is a dangerous exercise. Whatever name is proposed, there will always be some one to claim it should not be used as it is supposed to have some reserved meaning. To be on the safe side, we could propose to call '#%£*@!' what quantified uncertainty represents. For simplicity sake, we will nevertheless use the word 'support' for it, realizing that some authors limits 'support' to 'rational degree of belief'. The concept of support is used quite loosely, and worlds like 'weight of opinion', 'degree of belief'... might be more appropriate. In any case, the weights given by the quantified representation of uncertainty will be called 'degrees of support', or 'support' for short.

The term 'support' seems to be quite neutral when it comes to distinguish between objective and subjective supports. The concepts of chance, of objective probability, of propensities, of objective possibility... are covered by objective supports, whereas those of subjective probability, of subjective possibility, of belief, of credibility, of strength of opinion, of necessary commitment... are covered by subjective supports.

We only consider the support-belief induced by the available information, without regard to the source of this information. Gebhardt and Kruse (1993, and this volume), in their Context Model, have studied the importance of considering the nature of the source of information, and were able to provide a unified model that underlies probability theory, belief function theory and possibility theory. Their approach shares strong similarity with Dempster's approach (see section 7.1). In this paper we focus only on the representation of the resulting measure of uncertainty, leaving aside their source.

2.1. Objective versus subjective support.

A first distinction has just been introduced: objective (or physical) versus subjective (or epistemical) support.

Subjective support concerns 'probable opinion', 'belief'. What is meant by 'belief' in philosophy is left aside. What we mean here is the same epistemic concept as the one considered by the Bayesian probabilists. The value .7 encountered in the statement 'the probability of A is .7' is supposed to quantified some one's 'belief'. We use 'belief' in this non categorical sense. This belief is held by an agent at a time t. So we introduce the agent, and we will call it 'You' to enhance its human nature (an old suggestion made by De Finetti (1974, vol. 1, pg. 21)). That beliefs can be held by other beings can be argued, but such a discussion is not relevant to the presentation. We further assume that the agent, You, is unique, and that You are an ideal rational agent as we focus only on normative theories for uncertainty, not on descriptive or prescriptive theories. Descriptive theories would focus on how human beings really behave when faced with uncertainty, and prescriptive theories would tell how human beings must behave when faced with uncertainty. Normative theories, as considered here, are more neutral than the prescriptive ones. They just develop rules and constraints that should be satisfied by an 'idealized' agent, without regard to what is done nor should be done by a real agent.

The existence of an **objective support** is less obvious. The discussion about its existence is the same as the one about the concept of objective probabilities. Extremists like de Finetti claim 'probability does not exist' (1974). At the other extreme, positivists will only accept objective probabilities. In order to study objective support, we restrict ourselves to objective probability as it is the most common form of 'support'. We will try to be 'objective' in our presentation, but we acknowledge our personal bias toward the subjectivist approach.

2.2. Two types of measures.

When we state that the probability of something is .7, or its belief, its possibility... what kind of measures are we considering? Such measures could quantify two things: either the strength of Your opinion about which world in a set of possible worlds corresponds to the actual world, or just a measure about sets similar to the height, weight etc... We introduce these two types of measures: the opinion measure and the set measure.

a) Opinion Measures. In the first case, we call it ‘opinion measure’, we assume the existence of an actual world denoted ω_0 . You know that ω_0 can be any world in a set of possible worlds Ω. The measures of probability, belief, possibility etc.... quantify the strength of Your opinion about which world in Ω is the actual world ω_0. So the statement ‘$\text{Prob}(\omega_0 \in A \mid \omega_0 \in B) = .6$’ for $A,B \subseteq \Omega$ expresses the fact that the strength of Your opinion that ω_0 belongs to the set A given You know ω_0 belongs to the set B is quantified at a level .6, and so is it with the other measures of opinion we are going to analyze.

b) Set Measures. The second case, we call it ‘set measure’, is illustrated by the concept of proportions, of relative frequencies. Suppose a set of n individuals that satisfy property B among which r satisfy property A. The proportion of individuals Prop(A|B) that satisfies A among those who satisfy B is just r/n. As such, this ratio does not deserve to be called a probability, it is just an proportion. A probability is a property given to ‘an individual’ or ‘this individual’, a proportion is a descriptive property of a set. The relation between the proportion of A’s among the B’s and the probability that a B is an A is not immediate, and could only be achieved through the introduction of extra assumptions, like some equi-probability in being selected... Distinguishing between proportions and probabilities is important as many errors result from their confusion.

Proportions are evidently not subjective probabilities. Nevertheless, it seems natural to assume that their knowledge should influence Your belief that a B is an A. As the next examples illustrate it, even that link is not obvious.

Example 1: Coin and tack tossing.

Consider an agent who tosses a coin twelve times and who tosses also a tack twelve times. It happens the coin lands 9 times on heads and the tack ends 9 times point up. So both ‘heads’ and ‘point up’ occur with the proportion 3/4. What would be Your opinion about the outcome of the next throw (assumed to be performed under identical condition). In most cases, we can expect You would claim that the probability of heads on the next coin-experiment would be 0.5 whereas You would probably accept that the probability of point up on the next tack-experiment would be close to 0.75. Why such a difference? Because You have a strong a priori opinion about the probability that a coin falls heads. You would need much more than 12 results to change Your mind. When it comes to the tacks, Your a priori opinion is very weak. You can only use the observation to make up Your mind and You use the observed proportion as a best guess for Your probability. This example shows just that the relation between probability and proportion is hardly immediate and is influenced by many extra supplementary assumptions, in this case the a priori opinion that is essentially subjective.

Example 2: Urn with non equiprobable balls.

Considering an urn with 30 white balls (W) and 70 black balls (B). So the ratio 0.30 represent the proportion of white balls in the urn. A ball is going to be extracted from such an urn. What is Your opinion that the ball will be white? So far, this opinion can be anything, because nothing has been said on how the ball is going to be extracted. To assume by default that in case of ignorance every ball has the same chance of being selected is just a wishful thinking. The selection procedure must be explicitly stated, and if indeed every ball has the same chance to be selected, then You are entitled to accept that the probability that the next ball will be white is 0.30. Is this 0.30 still an objective property, like it was the case with the proportion 0.30? It depends only on the nature You would give to the concept of 'equiprobability' that underlies the selection procedure just described. If You assume that this equiprobability is objective, than the 0.30 is an objective probability. If You claim that the equiprobability is subjective, than the 0.30 is a subjective probability. So this example does not provide any help to decide about what is objective and what is subjective in probability theory. Even though the proportion is objective, the probability induced by it is not necessarily so.

c) Links. Thus the links between proportions and subjective probabilities is not obvious. What about their links with objective probabilities? And what are objective probabilities? Do they exist? Can we speak about the objective probability that heads will occur when tossing *a* coin or when tossing *this* coin? Can we state in a meaningful way that the objective probability of heads is 0.5? Surprisingly, the matter is hardly settled (Pollock, 1990). Theories for objective probabilities have been identified with finite frequency theories, then with limiting frequency theories, finally with hypothetical frequency theories. The subtlety of the last form is that we accept even 'virtual sequences' of physically possible events, and we are not restricted to actual events. These frequency theories concern '**indefinite probabilities**', i.e., probabilities for classes of individuals. Instead '**definite probabilities**' concern specific individuals. The difference between indefinite and definite probabilities is illustrated by 'the probability that heads occurs when tossing *a* coin' versus 'the probability that heads occurs when tossing *this* particular coin'. Under what conditions can we use the indefinite probability that would characterize 'a coin' when we want to determine the definite probability for 'this coin'? Today, this problem of 'direct inference' is yet not solved.

Hacking (1965) proposes to define objective probabilities, that he calls 'chances', as an indefinite probability . He claims that the 0.5 chance of observing 'heads' is a property of a 'chance setup'. The 0.5 is neither a property of the coin itself nor of the agent who observes the experiment, but of the whole physical setup, the coin, the tossing device, the physical environment on which the experiment is run...... The 0.5 could exist even in the absence of any agent. Strict subjectivists would defend that the 0.5 is not a property of the chance setup but only of the agent, You, who observes the experiment. They would defend strict determinism for what concerns the outcome of the experiment and the probability appears only in Your mind. Because of Your limited understanding, You cannot determine the outcome and You can only express Your belief about which outcome will prevail.

d) Propensities. The failure to define the concept of objective probabilities led to the introduction of the propensity theories. Initially, Popper (1959b) introduces propensities to be properties of individuals, hence propensity is equated to definite probability. The propensity is defined to be ‘the strength of the dispositional tendency for an experimental setup to produce a particular result on its singular trial’ (Fetzer, 1971). Mellor (1969) takes propensities to be ‘dispositions of objects to display chance distributions in certain circumstances’. As Pollock (1990) shows, these propensity theories can be generalized exactly like the objective probabilities theories, and all what has been said for the last hold for the first. So propensities and objective probabilities share the same difficulties when they must be defined. Hence Pollock (1990, page 37) goes as far as to introduce ‘nomic probabilities’ as a primitive concept that cannot be defined. He considers that nomic probability should nevertheless be analyzed in terms of their roles in reasoning. Their analysis must focus on how to assess them, how to manipulate them, and how to use them, neglecting finding their definition.

The conclusion of this analysis is that we failed to prove that objective probabilities exist. We cannot define them, and must accept them as primitive concepts. This failure is really feeding the skepticism of the subjectivists about the existence of any objective probability. Let us nevertheless accept that they exist, whatever they are. For simplicity sake, we will not anymore distinguish between objective and subjective supports. We will systematically speak about the support given by You at time t about the fact that a proposition is true or false. We acknowledge that our presentation has a strong subjectivist flavor and we often replace the word ‘support’ by ‘belief’. Nevertheless the presentation could be transformed into an objective context by considering that You, the agent, is some kind of robot ignorant of any subjectivity.

2.3. The evidential corpus.

The strength of the support-belief given by You at time t to the fact that a given proposition is true is defined relatively to a given evidential corpus, denoted EC_t^Y, i.e., the set made of the pieces of evidence in Your mind at time t. The evidential corpus EC_t^Y corresponds to Your background knowledge, to ‘all that You knows at t’. EC_t^Y is the set made of the pieces of information used by You at t to build Your beliefs.

For what concerns the construction of Your beliefs, EC_t^Y is composed of propositions accepted to be true by You at t. Some propositions bear directly on the domain on which Your beliefs are built. Other propositions translate rationality principle that Your beliefs should satisfy. For example, if You are a real Bayesian, You would put in EC_t^Y the list of possible events on which Your probabilities will be defined, Your opinion about which event will prevail, and some rationality requirements like the one that states that the probability given to two mutually exclusive events should be the sum of the probabilities given to the individual events, etc... Given EC_t^Y, You will assign a probability to every event.

One could claim that EC_t^Y uniquely determines these probabilities, a reminiscence of Carnap logical probabilities. Nevertheless, the derived probabilities can also be subjective as EC_t^Y can also contain Your personal opinions at t. It is hard to differentiate between subjective probabilities and logical probabilities induced by propositions that describe personal opinions.

The belief set of Gärdenfors (1988) is a subset of EC_t^Y, it is the set of propositions in EC_t^Y that induce the list of possible events on which beliefs will be distributed.

We do not include the beliefs assigned to the various events in EC_t^Y as we want to use it just as a description of the background from which You builds Your beliefs at t. If we had put the values of Your beliefs in EC_t^Y, then the beliefs induced by EC_t^Y would be nothing but those included in EC_t^Y. We use EC_t^Y as a ‘background’ and we say ‘EC_t^Y induces beliefs so and so’ just as we would say ‘under such a background, beliefs are so and so’.

2.4. Formalizing the domain of support.

We formalize the domain on which degrees of support are assessed.

a) Possible worlds. Our presentation is based on possible worlds (Bradley and Swartz, 1979) and support-belief will be given to sets of worlds. These sets of worlds, called ‘events’ in probability theory, will be elements of an algebra of sets.

Let L be a finite propositional language, supplemented by the tautology $\top$ and the contradiction $\bot$. Let Ω_L be the set of worlds that correspond to the interpretations of L and built so that every world corresponds to a different interpretation. Propositions identify the subsets of Ω_L, and the subsets of Ω_L denote propositions. For any proposition p in L, let $[\![p]\!] \subseteq \Omega_L$ be the set of worlds identified by p (i.e., those worlds where p is true).

We assume that among the worlds of Ω_L a particular one, denoted ω_0, corresponds to the actual world. Suppose the available information about the actual world ω_0 is imprecise, so it does not define ω_0 uniquely. All that can be stated about which of the possible worlds is the actual world ω_0 is the strength of the support (given by the agent at time t) that ω_0 belongs to that or that subset of Ω_L. We denote this degree of support by $S(\omega_0 \in A)$, or S(A) for short, where $A \subseteq \Omega_L$.

b) The frame of discernment. By definition the actual world ω_0 is an element of Ω_L. Let the frame on which Your uncertainty is defined be called the frame of discernment. Is Ω_L really the frame of discernment? We explore the nature of Ω_L and show that often Ω_L is not that frame of discernment.

The worlds of Ω_L are built logically by taking all the possible interpretations of L. Up to here, no epistemic constraints are yet introduced. So if L = {H, T, E} with the meaning that H (heads), T (tails) and E (exploding) are three outcomes of the coin tossing experiment (Exploding means that the coin disintegrates in the air before touching the ground, a favorite outcome for some philosophers). Then Ω_L contains eight worlds listed in table 1 and the Boolean algebra 2^{Ω_L} built from it contains $2^8 = 256$ propositions.

world	interpretation	only one of H,T,E	Ω	E impossible	EC_t^Y
ω_1 :	H & T & E	inconceivable			
ω_2 :	H & T & ¬E	inconceivable			
ω_3 :	H & ¬T & E	inconceivable			
ω_4 :	H & ¬T & ¬E	conceivable	x	impossible	x
ω_5 :	¬H & T & E	inconceivable			
ω_6 :	¬H & T & ¬E	conceivable	x	possible	x
ω_7 :	¬H & ¬T & E	conceivable	x	impossible	
ω_8 :	¬H & ¬T & ¬E	inconceivable			

Table 1: List of the different possible worlds built from L = {H, T, E}.

Then You learn the epistemic constraints that one and only one of H, T and E can occur: so there are only three worlds that are epistemically conceivable for You: those in $\Omega = \{\omega_4, \omega_6, \omega_7\}$. Given Ω, You can build another Boolean algebra 2^Ω that contains $2^3 = 8$ propositions. This new algebra is not a subalgebra of the first one as they do not share the same top element.

It must be realized that Ω_L might be itself inadequately limited. For example, You did not consider in L a proposition V corresponding to 'the coin violates the gravity laws and moves up for ever'. There is of course no way to recover such inadequate restriction of L, except by changing L itself. So L should be 'rich' enough to cover every possible outcomes, but this is of course just wishful thinking. In practice, L is what You are able to think of: it is doubtful that You would ever think to put V in L.

Note the concept of negation must be handled with care. Negation is relative to an algebra. So ¬H relative to Ω_L is $\{\omega_5, \omega_6, \omega_7, \omega_8\}$, whereas ¬H relative to Ω is $\{\omega_6, \omega_7\}$.

Formally, because of Your epistemical understanding of the propositions in L, some of the worlds of Ω_L are not conceivable to You at t. Let $\Omega \subseteq \Omega_L$ be the set of worlds conceivable by You at t given Your evidential corpus EC_t^Y. The set Ω is called the frame of discernment.

Of course EC_t^Y can say more about Ω. It can tell that some worlds in Ω are in fact considered as impossible by You at t (so it is the case for the worlds ω_1, ω_3, ω_5 and ω_7 in the coin tossing experiment if You consider E as impossible). Let $[\![EC_t^Y]\!]$ denote the set of worlds in Ω where all the propositions deduced on L from EC_t^Y are true (the worlds ω_4 and ω_6). Hence Your beliefs are essentially defined on $[\![EC_t^Y]\!]$. Nevertheless we can innocuously extend the domain of Your beliefs to Ω. So by construction, $[\![EC_t^Y]\!] \subseteq \Omega$. The worlds in Ω and not in $[\![EC_t^Y]\!]$ are considered as impossible to You at t, the worlds in Ω_L and not in Ω are inconceivable to You at t: impossible and inconceivable worlds should not be confused.

Your beliefs about ω_0 can only be expressed for the element of the Boolean algebra built from the worlds in Ω. The idea of speaking about the belief given by You to a set of worlds inconceivable to You seems difficult to accept and is thus rejected. Note that nothing requires ω_0 to be in Ω: the actual world can be 'inconceivable' to You at t, a situation usually not handled by the numerical representation and that could lead to inconsistency.

When $\Omega \neq \Omega_L$, it could be tempting to consider the set η of worlds of Ω_L not in Ω, to extend Ω by adding η to it, and so to define Your beliefs on Ω_L. We prefer to avoid this procedure as we feel that You could be in a state of beliefs where You can only express Your beliefs over the subsets of Ω. Creating the extra set η works innocuously in probability theory, and is thus usually introduced in such a context in order to avoid the difference between Ω and Ω_L. The introduction of the extra set η does not work innocuously with more general theories where the degree of support given to $A \cup \eta$ for $A \subseteq \Omega$ is not just the sum of the degrees of support given to A and to η. In these more general theories, if we add the extra set η, You would have to specifically assess Your beliefs for the subsets $A \cup \eta$ for all $A \subseteq \Omega$. This is not a realistic requirement as You do not know what these subsets represent as those worlds in η are by definition 'inconceivable' for You at time t.

To illustrate a case where $\Omega \neq \Omega_L$, consider a diagnostic procedure. You establish a list of mutually exclusive and hopefully exhaustive diseases. Given the patient's symptoms, You will create a belief about which disease prevails. But the list of diseases might be non exhaustive, some potential diseases might have been omitted, as would be the case for all those diseases still unknown today by You. How could You build a belief that the patient suffers from a 'still unknown disease', or from 'disease A or a still unknown disease'? Requiring that $\Omega = \Omega_L$ implies that such beliefs must be assessed, what we feel is not a reasonable requirement. Hence we prefer to accept that Ω and Ω_L might be different.

It is worth noticing that extending the belief domain from $[\![EC_t^Y]\!]$ to Ω was accepted, whereas extending it from Ω to Ω_L was not. The reason for such asymmetry is that, in the first case, You know that the worlds in Ω not in $[\![EC_t^Y]\!]$ are impossible, whereas You have no opinion about the worlds in Ω_L not in Ω. The distinction between those worlds in Ω and those in Ω_L not in Ω is also defined when considering the concept of 'awareness' (Fagin et al., 1995). Before expressing any opinion about a world, You must be 'aware' of its existence. You are aware of the worlds in Ω and You are not aware of the worlds in Ω_L not in Ω.

An extension of the concept of the frame of discernment has been proposed by Besnard et al. (1996), Jaouen (1997) and Perin (1997). They consider that the adequate frame is the distributive lattice built from the propositions in L. Inconceivable propositions are distinguished. Even though You know that H, T and E are pairwise inconsistent, H&T&E is different from both H&T and H&(T∨E), and H is different from H∨(T&E)... all these propositions belonging to the lattice. These distinctions cannot be described in classical logic based on Boolean algebras. Thanks to their generalization, the authors can solve the following problem. Suppose two sensors looking at a falling object. You know it can be a bomb (b) or a decoy (d), it can be large (l) or small (s), and that b and d are contradictory, and l and s are also contradictory. One of the sensor states that the object is b&l, the other that it is b&s. In classical logic, their conjunction is a contradiction (b&l&b&s = ⊥). Common sense might prefer the conclusion that 'it is a bomb, and the sensors disagree for the size'. Abandoning the Boolean algebras and using distributive lattices permits to solve this problem adequately. We will not further explore this promising generalization in this presentation.

c) Doxastic equivalence.

In the propositional language L, two propositions are logically equivalent iff the sets of worlds that denote them are equal. Beside this logical equivalence, there is another form of equivalence that concerns Your beliefs. Suppose You want to decide whether to go to a movie or stay at home tonight. You have decided to toss a coin, and if it is heads, You will go to the movie, and if it is tails, You will stay at home. (These are the pieces of evidence in EC_t^Y). Then 'heads' and 'going to the movie' are 'equivalent' from Your point of view as they share the same truth status given what You know at t. Of course, they are not logically equivalent (Kyburg, 1987). We call them doxastically equivalent (from doxa = an opinion, in Greek). Logical equivalence implies doxastic equivalence, not the reverse.

Definition: Two propositions p and q defined on L are doxastically equivalent (for You at t, i.e., given EC_t^Y) iff the sets of worlds $[\![p]\!]$ and $[\![q]\!]$, both subsets of Ω_L, that denote them share the same worlds among those in $[\![EC_t^Y]\!]$, i.e., $[\![EC_t^Y]\!] \cap [\![p]\!] = [\![EC_t^Y]\!] \cap [\![q]\!]$.

The importance of the doxastic equivalence is that any theory for representing uncertainty should give the same degree of support to two doxastically equivalent propositions. This constraint implies that the degree of support should not depend on the language used to express the propositions, a fact that is not necessarily true when studying human behaviors (Gigerenzer, 1996)

d) Complement. For $A \subseteq \Omega$, $\bar{A}$ denotes the set of worlds in Ω not in A. By definition, $A \cup \bar{A} = \Omega$.

e) The support relevant algebra. Whenever there is some support that ω_0 belongs to a set A, and to a set B, there is also a support that ω_0 belongs to their complement (relative to Ω), union and intersection. Therefore, the domain of the supports is assumed to be a Boolean algebra of subsets of the frame of discernment Ω (thus closed under union, intersection, complement, and containing Ω and Ø).

Let $\Re$ denote such a Boolean algebra of subsets of Ω on which some given support is built. Related to this algebra $\Re$, there exists a partition of Ω so that each element of $\Re$ is the union of elements of the partition. We call 'atoms' the elements of the partition of Ω related to $\Re$. We denote by $At(\Re)$ the set of atoms of the algebra $\Re$. When $\Re$ is the power set of Ω, the atoms of $\Re$ are the singletons of Ω.

Why do we introduce the credibility domain $\Re$, restricting Your beliefs to it, and we just do not accept that $\Re$ is the power set of Ω? The reason is that the propositional language L can be very rich, therefore the worlds of Ω can denote very precise propositions, and due to Your limited understanding or Your limited interest, You cannot express Your beliefs on such a detailed domain. When You want to assess Your beliefs about tomorrow weather in Brussels, You will not assess Your beliefs on the weather at every location on the Earth. You will restrict Yourself to Brussels even though L could be: {'Brussels weather is fine', 'New York weather is fine', 'Tokyo weather is fine', ...}. When asked about Your belief about Brussels weather, You build a credibility domain $\Re$ with two atoms: one where Brussels weather is fine, and one where Brussels weather is not fine. You will not build an atom where simultaneously Brussels weather is fine and New York weather is not fine and Tokyo weather is fine, etc... You just do not care about such a refined domain.

f) The support state. We have reached a point where we can formally define the support state of the agent You at time t. It is the quadruple $(\Omega, \Re, S, EC_t^Y)$ where Ω is the frame of discernment, $\Re$ is the Boolean algebra on which Your supports are assessed, S is a function of $\Re$ to [0, 1] that assigns to the elements of $\Re$ the measure of support given by You at t that the actual world belongs to these elements, and EC_t^Y is Your evidential corpus at t. Replacing S by a probability, a possibility or a belief function, we get the probability theory, the possibility theory or the transferable belief model, respectively. Support states as defined do not covered the theories where supports are defined as family of probability functions. To obtain these, the function S has to be appropriately adapted.

3. Generic and factual knowledge

Beside describing the major mathematical structure of a theory for the representation of uncertainty, we want to present the difference between generic and factual knowledge (Dubois and Prade, 1992, see also Pearl, 1990b, pages 384-5, who calls them knowledge and evidence, respectively). Such a discussion is rarely made as both types of knowledge are treated identically within the classical Bayesian theory, even though they fit somehow with the distinction we made between indefinite and definite probabilities in section 2.2. Their divergence becomes very apparent in the upper and lower probability theory when conditioning is analyzed. So we will focus on that special case.

Generic knowledge is a property of the population. A particular element of that population is observed and factual knowledge will be a property of that particular individual. Generic knowledge is a property of the set of worlds in Ω whereas factual knowledge is a property of ω_0 itself.

In predicate logic, generic knowledge is represented by quantified expression: ‘it exists...’ and ‘for all...’, whereas factual knowledge is represented by grounded propositions. In propositional logic, both types of knowledge are represented by propositions, the language being not sufficiently expressive to show the difference. Classical probability theory being built atop of propositional logic, the distinction is also usually hidden. Nevertheless we introduce the next example in order to illustrate the distinction.

Example 3. The killer’s sex.
Suppose John and Lucy are the only suspects for a murder case. A witness might have seen the murderer with a probability .6 and the person he saw was a male. Besides, You know that in the population, half the murderers are male. What would be the strength of Your opinion that the killer is John. A plain probabilistic analyses will lead to:

Prob(killer = John) =
    Prob(killer = John | witness saw the killer) Prob(witness saw the killer)
 + Prob(killer = John | witness didn’t see the killer) Prob(witness didn’t see the killer)
 = 1.00 x 0.60 + 0.50 x 0.40
 = 0.60 + 0.20
 = 0.80.

The computed probability measures of course the strength of an opinion. The 'actual world' corresponds to the name of the killer. Two pieces of knowledge are used in order to compute the 0.80. The term 0.60 results from a factual knowledge as it concerns the killer itself. The term 0.20 results from the combination of a factual knowledge (the 0.40

probability) and of a generic knowledge (the 0.50 probability that concerns the population of killers) that You project on the particular case. The combination of these two terms that results in the 0.80 value is typical of the Bayesian approach and fits perfectly well with the betting quotient interpretation given to a subjective probability (see section 6.5). Nevertheless using the base rate (the 0.50 probability) at the factual level can be criticized. In Court, the 0.80 probability would not be accepted as the generic information (the base rate) may not be introduced when trying to prove the culpability of John. Only the factual knowledge (the witness information) could be introduced and the probability a judge would accept is the 0.60. This allocation does not allow You to give the complement 0.40 to Lucy. The 0.40 may not be understood as a probability that the killer is Lucy.

In the transferable belief model, we will indeed consider that the belief that the killer is John is 0.60 whereas the belief that the killer is Lucy is 0.00. The evidentiary value model (Gärdenfors et al., 1983) would also claim that the belief that the killer is John is 0.60, where the 0.60 means essentially that 'the killer is John' is proved at that level. The 0.40 is not given to Lucy, but to ‘John is not proved to be the killer’, what does not mean that Lucy is proved to be the killer. The probability of provability approach would also give the same values (see section 8.3).

Example 4: Imprecise Database.

Suppose a database with seven cases and one field containing the age (table 2). It happens that the ages are imprecisely known. The age can be interval valued like for case 1 whose age is in the interval [13-17], or disjunctive like for case 2 whose age is either 15 or 40. Because of the imprecision, some proportions cannot be exactly known. Depending on the values given to the actual ages of the seven cases, we can construct a family of functions giving the proportion of cases falling in every subset of the age domain Ω = [0, 100] and compatible with the data.

				Case is <25				All cases <25			
		20-30		20-25	0-19	20-25	0-19			20-30	
Case	Age	Nec	Pos	Nec	Pos	Pos	Nec	Age	Incl	Nec	Pos
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1	13-17				1		1	13-17	1		
2	15 or 40				1			15	1		
3	11 or 50				1			11	1		
4	27	1	1					-			
5	21-23	1	1	1	1	1	1	21-23	1	1	1
6	23 or 45		1			1	1	23	1	1	1
7	15-23		1		1	1	1	15-23	1		1
Total		2	4	1	5	3	4		6	2	3

Table 2: Value of the age of seven individuals.

Suppose the interval [20, 30] of Ω. In order to compute the minimal and maximal proportions of cases that belong to [20, 30], we determine those cases that must belong to the interval whatever the values of their actual age, and those who might belong to it. In table 2, the columns Nec and Pos (columns 3 and 4) indicate those two groups of cases. Cases 4 and 5 belong to [20, 30] whatever the actual ages of these two cases. Cases 4 to 7 might belong to it as it is possible that the actual ages of these cases are in the interval [20, 30].

Let these two proportions be called the upper and lower proportions, denoted Prop*([20-30]) and Prop_*([20-30]), respectively. We have:

$\text{Prop}_*([20\text{-}30]) = 2/7$

$\text{Prop}^*([20\text{-}30]) = 4/7.$

Suppose an individual will be selected from the database and the selection procedure is such that every individual has the same chance of being selected. Now we can speak of probabilities: they result from the selection procedure, and the probabilities happen to be equal to the proportions because of the equiprobability of being selected. What would be the probability You would give to the fact that the age of the selected individual, denoted ω_0, would fall in the interval [20-30]?

Due to the imprecision of the proportions that results from the imprecision in the data, we can only build the family of probability functions compatible with the available data. By construction, this family is in one-to-one correspondence to the family of proportion functions compatible with the data. Let P denote the family of probability functions Prob define over Ω and compatible with the data. For any subset of Ω, we can only determine the upper and lower probabilities, denoted Prob* and Prob_*, respectively, by taking the extremes values given to that subset when Prob is constrained within P. We have:

$\text{Prob}_*([20\text{-}30]) = \min \{ \text{Prob}([20\text{-}30]) : \text{Prob} \in P\} = 2/7$

$\text{Prob}^*([20\text{-}30]) = \max \{ \text{Prob}([20\text{-}30]) : \text{Prob} \in P\} = 4/7$

These probabilities are numerically equal to the corresponding proportions because of the equi-probability of the sampling procedure. If the chance of being selected had depended on the individual, this direct relation between the probabilities and the proportions would disappear.

Now suppose You receive the information that the selected individual happens to be younger than 25 years old. This is a factual knowledge as it tells something about ω_0 itself. Among other it tells that the case 4 could not have been the one selected. Depending on the values given to the actual ages in agreement with the available data, we can compute the upper and lower values that Prob([20-30] | [0-25]) could achieve. This is done by taking every probability function Prob that belongs to P, computing Prob([20-30] | [0-25]) through the application of the Bayesian conditioning rule, and finding the extreme values these conditional probabilities could reach. It can be shown that the conditional probability that the age of the randomly selected individual is between 20 and 30 becomes:

$$\mathrm{Prob}_*([20\text{-}30] \mid [0\text{-}25]) = \frac{\mathrm{Prob}_*([20\text{-}25])}{\mathrm{Prob}_*([20\text{-}25]) + \mathrm{Prob}^*([0\text{-}19])} = 1/6$$

$$\mathrm{Prob}^*([20\text{-}30] \mid [0\text{-}25]) = \frac{\mathrm{Prob}^*([20\text{-}25])}{\mathrm{Prob}^*([20\text{-}25]) + \mathrm{Prob}_*([0\text{-}19])} = 3/7.$$

The columns 5 to 8 of table 2 provide the detail of the needed data to compute the upper and lower probabilities.

Let us now suppose another piece of information that states that none of the individuals in the database were older than 25. This is a generic knowledge as it concerns all the individuals in the database and not only the selected one. In that case, the database is transformed into a new data base, as given in columns 9 to 12 of table 2. For every case, the 'age' is obtained by intersecting the previous subset with the interval [0-25]. Case 2 was known to be either 15 or 40. As nobody was older than 25, we know now that case 2 is 15 years old, etc... For case 4, one might wonder why he was initially in the data base. Indeed the intersection is empty, he was known to be 27, and now we learn that everybody was younger that 25. There are two ways to handle that case. We can consider it as an error and eliminate it from the database, in which case only 6 cases are left over, all probabilities are normalized, and:

$$\mathrm{Prob}_*([20\text{-}30] \mid [0\text{-}25]) = 2/6$$

$$\mathrm{Prob}^*([20\text{-}30] \mid [0\text{-}25]) = 3/6.$$

We can also decide to keep it as an indication of some incoherence and not to normalize the data, keeping 1/7 as an amount of conflict, in which case:

$$\mathrm{Prob}_*([20\text{-}30] \mid [0\text{-}25]) = 2/7$$

$$\mathrm{Prob}^*([20\text{-}30] \mid [0\text{-}25]) = 3/7.$$

Notice that if the values of the database had been known precisely, not only the upper and lower probabilities would have been equal, but also the conditional probabilities that would result from both the factual and the generic knowledge would have been both equal to the conditional probabilities obtained by the application of the Bayesian rule of conditioning. This degenerescence explains why the distinction between generic and factual knowledge is not important in probability theory.

Example 5. Failure diagnosis.

Suppose an electrical equipment has failed and You know that one and only one circuit has failed. There are two types of circuits, the A- and the B-circuits made at the F_A and F_B factories, respectively. You also know that the A-circuits are always painted in green (G) and the B-circuits are always painted in white (W) and red (R) (Smets, 1997).

Two pieces of conditioning information are considered.

Case 1: Generic knowledge. You learn that all circuits made at factory F_B used in the failed equipment were painted white. Let EC_1 denote this piece of evidence. Given EC_1, B and W are doxasticaly equivalent. Indeed knowing that the circuit has been made at factory F_B is now equivalent to knowing that the circuit is white.

Case 2: Factual knowledge. You possess a fully reliable sensor that is only able to detect if the color of a circuit is red or not, so it cannot distinguish between green and white circuits. You learn that the sensor has reported that the broken circuit is not red. Let EC_2 denote that piece of evidence. Given EC_2, B and W are doxasticaly equivalent. Indeed knowing that the broken circuit has been made at factory F_B is now equivalent to knowing that the broken circuit is white.

The difference between these two doxastic equivalencies resides in the fact that the second concerns only the broken circuit, whereas the first concerns all circuits made at factory F_B. The way Your support is transformed by what You learn depends on the theory used to represent Your support (see Smets, 1997).

Example 6: Aircraft Recognition System.

Suppose a radar system built to recognize aircrafts. There are three types of aircrafts. For each of them, You know their flight characteristics and their shape. These pieces of information are generic. They are intrinsic part of Your detection system. Suppose You learn from Your spies that the enemy possesses only 3 airplanes of each type, that one of them seems to be unable to fly, that the weather is unfit for planes type 1, etc... The generic or factual nature of these new pieces of information is unclear, both characterizations can be defended. Suppose these pieces of information are valid and applicable for a whole day. They are generic for a systems working during that day. They are factual for a system working for the whole year. Suppose now that an airplane is detected. Its behavior and shape are obviously factual knowledge as they concern the specific airplane under consideration. This example was given to show that the distinction between generic and factual can be unclear and can be sometimes context dependent.

Example 7: A random process.

Suppose an urn that contains hundred balls: ten white balls (W), thirty red balls (R) and sixty black balls (B). A ball will be selected by a random procedure, and every ball has the same ‘chance’ to be selected. These pieces of information are generic. Your belief about the color of the selected ball is represented by the probabilities P(W) = .1, P(R) = .3, P(B) = .6. The ball is extracted, and You learn the ball is not red: this is a factual information that You use to revise Your belief that the ball is black. This is achieved by the application of the Bayesian conditioning rule, and Your belief that the ball is Black becomes 60/70. But You might also receive a generic information like the one that states that the probability that a red ball is selected is twice the probability that the other balls are selected. You will update the probabilities into the new probabilities: P’(W) = 10/130, P’(R) = 60/130, P’(B) = 60/130. This is of course not achieved by the Bayesian conditioning rule that translates the impact of a factual information. Generic information implies usually a complete revision of the initial probabilities, and is not described by any ‘all-purpose’ rule.

4. Theories for representing uncertainty.

Theories for imperfect data can be classified as symbolic-qualitative or numeric-quantitative theories. Symbolic theories are examined in another chapter of this handbook, and will not be studied here.

Before digging into the maze of the theories for uncertainty, a first remark deserves to be mentioned. It concerns the myopic attitude of most approaches to the representation of uncertainty. Authors very seriously argue about the meaning of words like probability, objective or subjective, disposition, propensity, corroboration, confirmation, belief, support, etc... But when it comes to propose a mathematical model, they miraculously agree in that they just accept the additivity axiom of probability theory, without hardly arguing about its adequacy. Let S(A) denote the support given to the event A, to the truth of the proposition A, etc... They just take it for granted that, when $A \cap B = \emptyset$, $S(A \cup B)$ must be the sum of S(A) and S(B), and of course they end up with probability functions. But why such a pervasive assumption? A correct approach would require a reasoned discussion as to how $S(A \cup B)$ compares with S(A) and S(B). So consider two events A and B with $S(A) > S(B) > S(\emptyset)$, and $A \cap B = \emptyset$. It must first be decided if $S(A \cup B)$ is decomposable, i.e., if $S(A \cup B)$ is a function of S(A) and S(B).

If it is, then it must be decided if either $S(A \cup B) = \max(S(A), S(B)) = S(A)$ or $S(A \cup B) > S(A)$. The first equality is proposed by Zadeh as justified for dispositional properties, and exploited in possibility theory. When strict inequality is justified, the additivity axiom of probability can sometimes be recovered after some appropriate monotone transformation of S insofar as it is granted that $S(A \cup B)$ is a regular function of S(A) and S(B). This cannot be achieved if $S(A \cup B) = \max(S(A), S(B))$. Then one must examine if $S(A \cup B) = S(A) + S(B)$ (or if a similar relation holds after applying some appropriate transformation to the S function) is justified or not. If it is, then probability theory might be appropriate. If not or worse if $S(A \cup B)$ is not decomposable, one should consider if belief/plausibility functions or some upper and lower probabilities theories are not more adequate. These comparisons do not allow us to derive uniquely the appropriate model, but at least it allows us to eliminate inadequate theories.

In any case, the additivity is too often uncritically accepted. Sometimes it even looks like the authors accept the additivity as a dogma, and never thought it could even be questioned. Such a blindness is kind of amazing, often encountered and very difficult to cure. Additivity should not be accepted as a default assumption, it has to be explicitly justified.

When comparing theories for representing imperfect data, it is important to consider both the static and the dynamic parts of the system, i.e., how imperfection is represented and how it evolves when new information is taken into consideration. Models can share the same mathematical structure at the static level, and their difference appears only once their dynamics is studied. For instance, confusion often occurred when theories based on belief functions are compared: Dempster-Shafer model, the transferable belief model, a family of upper and lower probability theories, random set theories and probability of provability theories all look identical at the static level. The difference appears once their dynamics is studied, in particular the conditioning process.

4.1. Modal logic.

The crudest form for representing uncertainty is based on modal logic. In modal logic, uncertainty is usually represented by the box (□) and diamond (◇) operators, where □A denotes ‘I know A’ or ‘I believe A’ or ‘A is necessary’ and ◇A denotes ‘I don’t know not A’ or ‘A is plausible’ or ‘A is possible’, depending if the modalities are used for epistemic, doxastic or ontic logics. We consider only the modality ‘I believe’. For what concerns Your belief about the truth status of a proposition p, we can say:
‘You believe p’ , in which case □p holds,
‘You believe ¬p’, in which case □¬p holds,
‘You don’t believe p and You don’t believe ¬p’, in which case ¬□p&¬□¬p holds.

We can thus represent three states of beliefs, the last representing ‘total ignorance’.

When studying quantitative theories for uncertainty, beliefs, etc..., it is always useful to study how the theories degenerate when their domains are limited to the two extreme values 0 and 1, instead of the whole range [0, 1] usually assumed. It will be shown that the theories based on probability functions degenerate into a modal logic where either □p or □¬p holds, i.e., a logic where □p = ¬□¬p. Such a degenerated probability function cannot express total ignorance. The other theories for quantified beliefs achieve that goal.

4.2. Generalizing the concept of set : fuzzy set theory.

Before studying the theories for imperfect data, we examine the concept of fuzzy sets.

Fuzzy sets.

Classically, sets are crisp in the sense that an element either belongs to a set or does not belong to it. Zadeh (1965) introduces the idea of non-crisp sets, called fuzzy sets. Fuzziness is related to the use of ill defined predicates like in 'John is tall'. The idea is that belonging to a set admits a degree that is not necessarily just 0 or 1 as it is the case in classical set theory. For some elements of the universe of discourse, we cannot definitively say that it belongs or not to the set. At most we can assess some degree of membership $\mu_A(x)$ of the element x to the fuzzy set A. This function generalizes the classical indicator function $I_A(x)$ of a set:

$I_A(x) = 1$ if $x \in A$
$I_A(x) = 0$ if $x \notin A$

Zadeh replaces the range {0, 1} by the interval [0, 1].

New concepts like fuzzy numbers (e.g. several, few), fuzzy probability (likely), fuzzy quantifiers (most), fuzzy predicates (tall), and the impact of linguistic hedges (very) can be formalized (Dubois and Prade, 1980). Classical set operators like union, intersection and negation have been generalized. The most classical solution is based on the min-max operators:

$\mu_{\overline{A}}(x) = 1 - \mu_A(x)$
$\mu_{A\cup B}(x) = \max (\mu_A(x) , \mu_B(x))$
$\mu_{A\cap B}(x) = \min (\mu_A(x) , \mu_B(x))$

Other operators have been proposed that belong to the family of triangular norms and co-norms (Dubois and Prade, 1985, Yager, 1991). The generalization of the implication operator turns out to be less obvious, especially when it is considered in the context of the modus ponens as encountered in approximate reasoning (Smets and Magrez, 1987, Smets, 1991b, Dubois and Prade, 1991XX).

The law of excluded middle does not always apply to fuzzy sets. Indeed $\mu_{A\cap\overline{A}}(x) = \min (\mu_A(x) , \mu_{\overline{A}}(x))$ can be larger than 0. This may look odd at first sight. It translates nothing but the fact that a person can be somehow tall and not tall simultaneously, a perfectly valid property.

Mathematically, fuzzy set theory generalizes the concept of set. This notion can be used wherever sets can be used, and therefore is not restricted to any particular form of imperfect data. Nevertheless its most common domain of application is for the modeling of weighted imprecise information. As we mentioned in the introduction, weighted imprecise information results from an ordering that represents the compatibility or the similarity between the information that characterizes the world and the imprecise information. The similarity relation induces a fuzzy set. The grade of membership given to a world expresses the intensity with which this world is compatible with the imprecise information. When I know that ‘Peter has a few children’, the value $\mu(\omega_3)$ given to world ω_3 where Peter has three children is equal to the compatibility of ‘having three children’ with ‘having a few children’, and similarly for the other possibilities.

Fuzzy set theory is not probability theory.
Several authors have tried to disregard fuzzy set theory by claiming that it is subsumed by probability theory. Fuzzy set theory concerns the belonging of a well-defined individual

to an ill-defined set whereas probability concerns the belonging of a not yet defined individual to a well-defined set. Some have tried to claim that the grade of membership of a man whose height is 1.80 m to the set of tall men is nothing but the probability that You qualified as ‘tall’ a person whose height is 1.80 m.

$\mu_{Tall}(1.80) = P(\text{You say ‘Tall’} \mid \text{height} = 1.80)$

One could defend such attitudes, but once conjunctions and disjunctions concepts are introduced, such solutions hardly resist. A membership function has some connections with a likelihood function, but this is only one possible view (see section 5.3.1).

There are mathematical relations between fuzzy set theory and probability theory but the problem is not with the mathematical comparison but with a comparison of the problems they try to modelize. Fuzziness concerns imprecision, probability has to do with uncertainty. Of course, imprecision induces uncertainty, hence fuzziness induces uncertainty. When I know that John is tall, I can build a probability measure on John’s height (for instance if I have to bet). This does not mean that the grade of membership is a probability (Smets, 1982a).

4.3. The Quantification of Uncertainty: Fuzzy Measures.

Sugeno (1977) introduced the concept of fuzzy measures that provides a general setting for the representation of uncertainty associated with a statement ‘ω belongs to A’ where A is a crisp set (generalization to a fuzzy set A is possible but not important here) and ω is a particular arbitrary element of Ω which is not *a priori* located in any of the subset of Ω. A fuzzy measure $g:2^{\Omega}\rightarrow[0, 1]$ defined on finite space Ω satisfies the following requirements:

G1: $g(\emptyset) = 0$ $g(\Omega) = 1$
G2: for all $A,B\subseteq\Omega$, if $A\subseteq B$, then $g(A)\leq g(B)$

Fuzzy measures for finite Ω are just normalized measures, monotone for inclusion. Even though they have been called ‘fuzzy’ measure, they should not be confused with fuzzy sets.

Given a fuzzy measure g, we can define its dual measure g’ by:

$g'(A) = 1 - g(\bar{A})$ for all $A\subseteq\Omega$.

It can be proved that g’ is also a fuzzy measure. In practice, today’s measures of uncertainty all satisfy the extra requirement that once $g(A) > g'(A)$ for some $A\subseteq\Omega$, then $g(B) \geq g'(B)$ for all $B\subseteq\Omega$. In such a case we propose to call g a potential support function and g’ a necessary support function, denoting them by g^* and g_*, respectively, and to speak of potential and necessary supports. The pairs possibility-necessity functions, plausibility-belief functions, upper-lower probability functions all satisfy the duality requirement, the first function or each pair is a potential support function, and the second is a necessary support function.

Probability functions are autodual in that $g(A) = 1 - g(\overline{A})$, hence $g^* = g_*$.

In general, in order to specify a fuzzy measure, a value must be given for every $A \subseteq \Omega$, hence if $|\Omega| = n$, then 2^n values must be provided. Simplification is achieved when the fuzzy measure is a decomposable measure (Dubois and Prade, 1982, Weber, 1984), i.e., when it satisfies relations like:

$g(A\&B) = g(A)*g(B) \qquad \text{if } A \cup B = \Omega$

or $g(A \vee B) = g(A)*g(B) \qquad \text{if } A \cap B = \emptyset$

and where * is some binary operator. In such a case, knowing g on n appropriately chosen subsets of Ω (the singletons or their complements) is sufficient to know g on every subset of Ω.

5. Possibility and Necessity Measures.

Possibility can be approached in at least three different ways: through modal logic, Baconian possibility or Zadehian possibilities. In modal logic, one can express that some proposition is possible and/or necessary, by using the $\Box$ and $\Diamond$ operators (Chellas, 1980). In Baconian possibility, also called ordinal (Schum, 1994), one considers only the possibility order among the propositions. In Zadehian possibility, also called cardinal, the values given to the degrees of possibility get some intrinsic meaning. We focus first on the Zadehian form, and show that the Zadehian and Baconian forms are essentially identical if one restrict oneself to min-max operators.

The difference between Baconian and Zadehian possibilities, or equivalently between ordinal and cardinal possibilities, is at the core of the difference between possibility theory à la Dubois and Prade (1997) and à la Zadeh (1978).

5.1. Possibility measure.

Categorical imprecise information such as ‘John's height is above 170’ implies that any height h above 170 is possible and any height equal or below 170 is impossible.

In modal logic, the proposition ‘John's height is above 155’ is possible and necessary in the face of the above piece of information, whereas the proposition ‘John's height is above 180’ is possible without being necessary, and finally the proposition ‘John's height is below 150’ is neither possible nor necessary. These modal expressions can also be represented in possibility theory, using the possibility values 1 for ‘possible’ and 0 for ‘impossible’, and the necessity values 1 for ‘necessary’ and 0 for ‘not necessary’.

It might happen that, given what You know, You might feel that some sets of values are more 'possible' than other. You are neither deciding nor betting on what the height is. You are only arguing about what the height might be. The origin of the ordering is usually due to the presence of some fuzziness in the underlying information You use. When the predicate itself is vague, like in 'John is tall', it is clear that possibility can admit degrees.

The question to know if graded possibilities can be defined when every piece of information You use is crisp is an open question. The following examples try to show that non fuzzy (crisp) events can admit different degrees of possibility, these appearing because some of the background knowledge is somehow fuzzy.

Example 8. Soft balls in a box.
Suppose a box in which You try to squeeze soft balls (Zadeh, 1978). You can say: it is possible to put 20 balls in it, impossible to put 30 balls, quite possible to put 24 balls, but not so possible to put 26 balls...These degrees of possibility are degrees of realizability (how difficult it is to squeeze the balls in the box), no fuzziness seems to be involved (except maybe on the definition of a soft ball) , but in any case these degrees are unrelated to any random process.

Example 9. Possible sales.
Suppose You ask a salesman about his forecast about next year sales. He could answer: it is possible to sell about 50KEcu, impossible to sell more than 100KEcu, quite possible to sell 70KEcu, hardly possible to sell more than 90KEcu... His statements express what are the possible values for next year sales. What the degrees of possibility express are essentially the strength of the opinion of the salesman about the sale capacity. Besides, the salesman could also express his beliefs about what he will actually sell next year, but this concerns another problem for which the theories of probability and belief functions are more adequate. Of course some links exists between the possibilities about next year sales and the beliefs about next year actual sale, just as we have developed the pignistic transformation that links beliefs to probabilities needed for decision making (see section 8.1). Unfortunately, the exact nature of that link is not yet settled.

The two last examples illustrates two forms of possibilities, a physical form and an epistemic form. The difference between the two forms can be recognized by their linguistic different uses: “it is possible for” and "it is possible that” (Hacking, 1975). In the first case, the possibilities expresses a physical property: one could have said “it is possible for 25 balls to fit into the box”. In the second case, the possibilities expresses an epistemic property: one could have said “it is possible that next year sales will be 100KEcu”. Physical possibilities are related to realizability, like in the soft balls (example 3) and Hans’ eggs next example (example 10). Epistemic possibilities are related to acceptance and epistemic entrenchment (Gärdenfors, 1988, Dubois and Prade, 1991)

5.2. Possibility functions.

Let $\Pi:2^{\Omega}\to[0, 1]$ be the possibility measure defined on a space Ω with $\Pi(A)$ for $A\subseteq\Omega$ being the degree of possibility that A (is true, occurs...). The fundamental axiom is that the possibility $\Pi(A\cup B)$ of the union of two sets A and B is the maximum of the possibility given to the individual sets $\Pi(A)$ and $\Pi(B)$. (Zadeh 1978, Dubois and Prade, 1988):

$$\Pi(A\cup B) = \max (\Pi(A) , \Pi(B)). \qquad (5.1)$$

Usually one requires also $\Pi(\Omega) = 1$, but this last requirement could easily be neglected as it is not really necessary.

In modal logic, the necessity of a set is the negation of the possibility of its complement. Identically, in possibility theory, one defines the necessity measure N(A) given to a set A by:

$$N(A) = 1 - \Pi(\overline{A})$$

In that case, one has the following:

$$N(A\cap B) = \min (N(A) , N(B))$$

Beware that one has only:

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

$$N(A\cup B) \geq \max (N(A) , N(B)),$$

equalities being only achieved in special cases. Whenever $\Pi(A\cap B) < \min(\Pi(A\cap\overline{B}), \Pi(\overline{A}\cap B))$, the inequality will be strict.

Related to the possibility measure $\Pi:2^{\Omega}\to[0, 1]$, one can define a possibility distribution $\pi:\Omega\to[0, 1]$, such that:

$$\pi(x) = \Pi(\{x\}) \qquad \text{for all } x\in\Omega.$$

Thanks to (5.1), one has

$$\Pi(A) = \max_{x\in A} \pi(x) \qquad \text{for all A in } \Omega.$$

Min-Max operators.

When only the max and min operators are used in possibility theory (and in fuzzy set theory), we benefit from a very nice and important property: the values given to the possibility measure or to the grade of membership are not intrinsically essential. Only the ordering they create among the elements of the domain is important. Any strictly monotonous transformation leaves the ordering unchanged. Therefore a change of scale will not affect conclusions. Under that conditions, the ordinal and cardinal forms of possibility theory turn out to be equivalent, and the distinction between Baconian and Zadehian possibilities might be neglected. As far as most applications of possibility theory are based on the min-max operators, the distinction is usually not mentioned. This similarity disappears once addition and multiplication are introduced.

Example 10. Hans' breakfast.

As an example of the use of possibility measure versus probability measure, consider the number of eggs that Hans is going to order tomorrow morning (Zadeh, 1978). Let $\pi(u)$ be the degree of ease with which Hans can eat u eggs. This degree is a measure of the extent to which his stomach, or his mood, can cope with a breakfast with u eggs. Let p(u) be the probability that Hans will eat u eggs at breakfast tomorrow. Given our knowledge, assume the values of $\pi(u)$ and p(u) are those of table 3.

u	1	2	3	4	5	6	7	8
$\pi(u)$	1	1	1	1	.8	.6	.4	.2
p(u)	.1	.8	.1	0	0	0	0	0

Table 3: The possibility and probability distributions associated with X.

We observe that, whereas the possibility that Hans may eat 3 eggs for breakfast is 1, the probability that he may do so might be quite small, e.g., 0.1. Thus, a high degree of possibility does not imply a high degree of probability, nor does a low degree of probability imply a low degree of possibility. However, if an event is impossible, it is bound to be improbable. This heuristic connection between possibilities and probabilities may be stated in the form of what might be called the **possibility/probability consistency principle** (Zadeh, 1978).

5.3. Possibility theory and other theories.

Possibility theory is neither the only nor the first theory proposed to quantify uncertainty where the main axiom is the maximum rule for the union,

$$\Pi(A \cup B) = \max(\Pi(A), \Pi(B)), \qquad (5.1)$$

or a transformation of it. Links exist with the likelihood functions, with Shackle's measures of potential surprise, with Spohn's measures of disbelief, with consonant plausibility functions (i.e., a plausibility function that satisfies the maximum rule).

5.3.1. Possibility theory and likelihood theory.

In probability theory, the concept of a likelihood function is often used (Edward, 1972). Let $P(x|\theta)$ be the conditional probability of an event $x \subseteq X$ given a parameter θ where

$\theta \in \Theta$. Suppose $P(x|\theta)$ is known for every $x \in X$, and every $\theta \in \Theta$. Suppose You observe some $x_0 \in X$, what is You belief about the value of θ? One solution is to define the so called likelihood function $l(\theta|x_0)$ over Θ with: $l(\theta|x_0) = P(x_0|\theta)$. The best 'supported', most 'likely' value for θ might be evaluated as the value of θ that maximizes the likelihood function, and this principle underlies a major part of statistical theory. If one tries to define a likelihood over the subsets of Θ, not only on its elements, a classical solution is to define the likelihood function into a function on 2^Θ, with $l(A|x_0) = \max_{\theta \in A} l(\theta|x_0)$ for $A \subseteq \Theta$.

The likelihood $l(\theta|x_0)$ can also be viewed as the degree of possibility that the actual value is θ. Thus $l(\theta|x_0)$ is assimilated to $\pi(\theta)$, and the generalization of the likelihood function to subsets of Θ is identical to what is obtained with possibility measures. Possibility and likelihood theories share common properties. These properties have not much been studied for historical reasons: likelihood theory concerns statistical inference whereas possibility theory concerns approximate reasoning (Smets, 1982b, Thomas, 1979, 1995, Dubois, Moral and Prade, 1995).

5.3.2. Shackle measure of potential surprise.

Shackle (1969) proposes the idea of a measure of potential surprise Surp that satisfies (5.1) when $Surp(A) = 1 - \Pi(A)$. Its motivations are identical to those that lead Zadeh (1978) to develop his possibility theory, but Shackle does not much elaborate on his approach from a formal point of view.

5.3.3. Spohn's measure of disbelief

Spohn (1990) proposes a measure of disbelief, denoted κ with $\kappa:\Re \rightarrow [0, \infty)$, the largest κ the largest the disbelief that A contains the actual world. His measure is a transformation of a consonant plausibility measure with $\kappa(A) = -\log(pl(A))$ (Dubois and Prade, 1991). Conditioning corresponds to Dempster's rule of conditioning (in its normalized form). Spohn also captures Jeffrey's rule of conditioning (Jeffrey, 1983), these results have been since generalized within the TBM (Smets, 1993d). In Spohn's model, Ø does not receive a well defined value, just as -log(0). Just as a consonant plausibility function is a possibility function, so is Spohn's measure up to the transformation. Nevertheless possibility functions and Spohn's measure diverge in the way conditioning is achieved, what just translates the fact that Dempster's rule of conditioning is not applicable in qualitative possibility theory.

5.3.4. Epsilon possibilities.

Adams (1975) proposes to define a set of propositions X to be p-consistent iff for any $\varepsilon > 0$, there exist a probability function P such that $P(A) \geq 1-\varepsilon$ for all A in X. This idea has been used by Pearl (1990a) in order to handle non-monotonic reasoning based on defaults rules, i.e., rules that admit exceptions. Let $\alpha \rightarrow \beta$ be the default rule that reads 'normally if α holds, then β holds'. Given a set Δ of default rules, he defines a class of probability functions $\mathcal{P}_\varepsilon$ so that for each P in $\mathcal{P}_\varepsilon$ and each default in the Δ, $P(\beta|\alpha)>1-\varepsilon$. His ε-calculus, and the so-called Z systems he proposes for default reasoning, has been the source of nice developments, but it happens that this theory does not require the concept of probability, that it is much more general that it seems at first sight. In fact, the ε-calculus creates a stratification of Δ, such that reasoning is performed by using all rules above the first level where some rules contained at that level contradict the previous deductions. The same results can be achieved when using possibility functions (Benferhat et al., 1992) or belief functions (Benferhat et al., 1995). In fact that last framework turns out to be the most general, what is not surprising as probability functions and possibility functions are just special cases of plausibility functions. Within this framework where all basic belief masses are close to zero except one that is close to one, it is possible to define what are the extra requirements that lead to the many theories that have been developed to handle default logic. That general framework is thus useful to compare the assumptions underlying the various approaches for default reasoning that have been proposed, sometime in a quite ad hoc way.

5.4. Relation between fuzziness and possibility.

Zadeh has introduced both the concept of fuzzy set (Zadeh, 1965) and the concept of possibility measure in its cardinal sense (Zadeh, 1978). The first allows one to describe the grade of membership of a well-known individual to an ill-defined set. The second allows one to describe what are the individuals that satisfy some ill-defined constraints or that belong to some ill-defined sets.

For instance $\mu_{\mathrm{Tall}}(h)$ quantifies the membership of a person with height h to the set of *Tall* men and $\pi_{\mathrm{Tall}}(h)$ quantifies the possibility that the height of a person is h given the person belongs to the set of *Tall* men. Zadeh postulates the following equality :

$$\pi_{Tall}(h) = \mu_{Tall}(h) \qquad \text{for all } h \in H$$

where H is the set of heights = $[0, \infty)$

The writing is often confusing. It states that the possibility that a tall man has a height h is equal *numerically* to the grade of membership of a man with height h to the set of tall men. It would have been better written as

$$\text{If } \mu(Tall|h) = x \text{ then } \pi(h|Tall) = x \qquad \text{for all } h \in H$$

This expression avoids the confusion between the two concepts. It shows that they share the same scale without implying that a possibility is a membership and *vice versa*. The previous expression clearly indicates the domain of the measure (sets for the grade of membership μ and height for the possibility distribution π) and the background knowledge (the height h for μ and the set *Tall* for π). The difference is analogous to the difference between a probability distribution $p(x|\theta)$ (the probability of the observation x given the hypothesis θ) and a likelihood function $l(\theta|x)$ (the likelihood of the hypothesis θ given the observation x).

6. Probability Theory.

The probability measure quantifies the degree of probability P(A) that an arbitrary element $X\in\Omega$ belongs to a well-defined subset $A\subseteq\Omega$. It satisfies the following property :

P1: $P(\emptyset) = 0 \qquad P(\Omega) = 1$
P2: For all $A, B\subseteq\Omega$, if $A\cap B=\emptyset$, $P(A\cup B) = P(A) + P(B)$
P3: For all $A, B\subseteq\Omega$, if $P(B)>0$, then $P(A|B) = P(A\cap B) / P(B)$

where P(A|B) is the probability of that $X\in A$ given it is known that $X\in B$. Such definition can be extended to fuzzy events (Zadeh, 1968, Smets 1982a) which further enhances, if still needed, the difference between probability and fuzziness. As an example consider the probability that the next man who enters the room is a tall man. Could we say that such a probability is .7 or is that probability itself a fuzzy probability? This distinction is still unresolved, what might explain today's lack of interest in the topic.

Related to the probability measure $P:2^{\Omega}\rightarrow[0, 1]$, one defines a probability distribution $p:\Omega\rightarrow[0, 1]$ such that :

$$p(x) = P(\{x\}) \qquad \text{for all } x\in\Omega.$$

By property P2,

$$P(A) = \sum_{x\in A} p(x) \qquad \text{for all } A\subseteq\Omega.$$

Notice that the relation between P and p is similar to the one between Π and π, (but not as the one between bel and m, as bel and m are both defined on the same frame 2^{Ω}, see section 8).

Since its beginning as a theory for uncertainty in the 17th century, probability has been given at least four different meanings.

6.1. The classical theory.

Laplace distinguishes between physical probability that he called 'degree of possibility' and epistemic probability that he called 'probability'. For that last case, he assumes the existence of a fundamental set of equipossible events. The probability of an event is then the ratio of the number of favorable cases to the number of all equally possible cases. Of course, the concept of 'equally possible cases' is hardly defined in general. It works with applications where symmetry can be evoked, as it is the case for most games of chance (dice, cards...). When symmetry cannot be applied, the Principle of Insufficient Reason is evoked (it is also called the Principle of Indifference (Keynes 1921)). It essentially states that alternatives are considered as equiprobable if there is no reason to expect or prefer any one over the other. As nice as it might seem, the Principle of Insufficient Reason is not acceptable as Bertrand's paradox kills it (Von Mises, 1957), and it is even a very dangerous tool whose application has led to many difficulties in probability theory. It is hardly defended today.

Example 11: Bertrand's paradox.
Given a bottle with a mixture of water and wine. All we know is that the mixture contains at least as much water as wine, and at most twice as much water as wine. Consider first the ratio of water to wine. It is between 1 and 2. So by the Principle of Insufficient Reason the probability of the ratio lying between 1 and 1.5 is 0.50. Consider next the ratio of wine to water. It is between 0.5 and 1. So by the Principle of Insufficient Reason the probability of the ratio lying between 2/3 and 1 is 2/3. These two results are not compatible, as the two probabilities concern the same event: the event 'the ratio water/wine is between 1 and 1.5' is equivalent to 'the ratio wine/water is between 2/3 and 1' They are doxastically equivalent, so they should share the same probabilities.

Example 12: Total ignorance.
Suppose E_1, E_2 and E_3 are the labels of three mutually exclusive and exhaustive events. You know nothing about what are the events described by the labels. All You know is their number. There is no reason whatsoever to give more support to any of them then to any other, so the three events share the same degree of support. Consider now the event $E_1 \cup E_2$, and compare it with E_3. Do You have any reason to give more support to $E_1 \cup E_2$ then to E_3. Remember that You do not know what the events are. So there is no more reason to support one than the other. So $E_1 \cup E_2$ and E_3 should receive the same degree of support. Of course, this violates probability axioms.

6.2. Relative frequency theory.

Some people claim that probability is essentially the limit to which converges the relative frequency under repeated independent trials (Reichenbach 1949, von Mises 1957). This definition is not concerned with capturing commonsensical notions. It tries to comply with the operationalist version of scientific positivism: theoretical concepts must be reducible to concrete operational terms. It is strongly related to the concept of proportion.

It is by far the most widely accepted definition even though it has been shown not to resist criticisms. Convergence limits cannot be observed, it does not apply to single events, it suffers from the difficulty of specifying the appropriate reference class, it never explains how long must be a long run that will converge to its limit... Nevertheless, it 'works' and this pragmatic argument explains its popularity.

6.3. Subjective (Bayesian, personal) probability.

For the Bayesian school of probability, the probability measure quantifies Your (You is the agent) belief that an event will occur, that a proposition is true. It is a subjective, personal measure.

The additivity of the probability measure (axiom P2) is essentially based on betting behavior arguments. Bayesians define the value p of P(A) as the 'fair price' \$p You propose that a player should pay to play a game against a banker, where the player receives \$1 if A occurs and \$0 if A does not occur. The concept of fairness is related to the fact that after deciding p You are ready to be either the player or the banker. In order to avoid a Dutch book (i.e., a set of simultaneous bets that would lead to a sure loss), You must assess the probability of the subset of Ω according to P1 and P2. The justification of P3 by diachronic[2] Dutch books (Jeffrey, 1988, Teller, 1973, 1976) is less convincing as it is based on a Temporal (Diachronic) Coherence postulate (Jeffrey, 1988, Earman, 1992) that can be objected to. It requires that the way You commit Yourself now to organize Your bets after A has occurred if it occurs should be the same as the bets You would accept once A has occurred. The Temporal Coherence claims that hypothetical bets (bets on the hypothesis that A occurs) should be equated to factual bets (bets after A has occurred) (Savage, 1954, De Finetti, 1974).

Another algebraic justification for the use of probability measure to quantify beliefs is based on Cox's axiom (Cox, 1946). It states essentially that the belief of $\bar{A}$ should be a function of the belief of A, and the belief of 'A and B' should be a function of the belief of A given B and the belief of B. Adding a strict monotonic requirement leads to the conclusion that the probability measure is the only measure that satisfies both requirements (Dubois et al., 1991, Paris, 1994). Cox's justification can nevertheless be criticized. Strict monotony kills possibility measures and the first requirement is of course not accepted in the theories based on possibility and belief functions (Clarke et al., 1991).

6.4. Logical probabilities.

Some attempts have been proposed to avoid the subjective component of the Bayesian probability. It fits with the objectivity one likes to defend for scientific rationalism.

Keynes (1921) defined probability as a logical relation between a proposition and a corpus of evidence. While propositions are ultimately either true or false (no fuzzy propositions are involved here), we express them as being probable in relation to our current knowledge. A proposition is probable with respect to a given body of evidence regardless of whether anyone thinks so. The logical probability P(p|q) is defined usually as the measure of the set of worlds where p&q hold divided by the measure of the set of worlds where q hold. The nature of this measure is unclear, only ad hoc solutions have been proposed.

---

[2] i.e., when time is involved, where there might be some bets before and some after a given event.

The concept of Corroboration introduced by Popper (1959a) and the concept of Confirmation introduced by Carnap (1950, 1952) both fit with the overall schema of defining a logical measure of probability. Bayesians accept the same kind of relation between knowledge and a proposition but admit it is subjective and therefore that the probability of a proposition is not an objective property that exists regardless of whether anyone thinks so.

This program unfortunately fails to explain how to define the probability weight to be given for these relations. On that point the strongest are the Bayesians who can use their betting behavior as a guideline on how to assess probabilities. The existence of such operational method to assess a measure of probability is important as it provides a meaning to the .7 encountered in the proposition " the probability of A is .7". The lack of such well-established and widely accepted operational meaning in fuzzy set theory, in possibility theory, in upper and lower probabilities theory, and in belief functions theory is the source of some weakness (see nevertheless Smets and Magrez (1988) for fuzzy set theory and Smets and Kennes (1994) for the transferable belief model). In qualitative possibility theory, the use of a qualitative possibility scale bypasses this problem.

6.5. The Dutch book argument.

The Dutch book argument provides a nice justification for using the probability model. A Dutch Book is a set of bets so organized that the player looses whatever event occurs. Given its importance, we illustrate its use to justify the additivity rule of probability functions.

Suppose three mutually exclusive and exhaustive events, A, B and C. I create three bets, denoted B1, B2 and B3, with:
Bet B1: the player wins 1$ if A, otherwise nothing.
Bet B2: the player wins 1$ if B, otherwise nothing.
Bet B3: the player wins 1$ if A or B, otherwise nothing.

You propose the prizes that should be paid by the player to enter each of the three bets. Let they be .2$, .5$ and .6$ respectively. Then I decide You are the player for B1 and B2, and the Banker for B3. Your gains are listed in table 4. If A occurs, You paid .2$ to enter B1 and get 1$, hence Your gain is .8$, You paid .5 to enter B2, hence Your gain is -.5$, and You received from the player .6$ to enter B3 and You pay him 1$, hence Your gain is -.4. So if A occurs, You total gain is .8 - .5 - .4 = -.1$. Similarly if B occurs, Your total gain is -.1$, and so is it also if C occurs. So whatever event occurs, You always loose .1$, hence a Dutch book has been made against You. The only way to avoid the Dutch book is by requiring that the price of B3 be the sum of the prices of B1 and B2, what means that the additivity rule of probability measure must be obeyed in order to avoid to become a 'money pump'.

If...occurs	B1	B2	B3	Your Gain
A	.8	-.5	-.4	-.1
B	-.2	.5	-.4	-.1
C	-.2	-.5	.6	-.1

Table 4: Your gain in each bet, and total gain, according to the event that occurs.

Similar arguments can be built to explain the conditioning rule. Conditional bets are defined where for instance You win if A, loose if B, and bet is canceled if C, in which case the banker gives back to the player the prize paid to enter the game. Behind the temporal coherence axiom, we find the assumption that the bets once the event C has occurred should be the same as the conditional bets.

7. Upper and lower probability theories.

Smith (1961, 1965), Good (1950, 1983) and Walley (1991) suggested that personal degrees of belief cannot be expressed by a single number but that one can only assess intervals that bound them. The interval is described by its boundaries called the upper and lower probabilities. Such interval can easily be obtained in a two-person situation when one person, Y_1, communicates the probability of some events in Ω to a second person, Y_2, by only saying that, for each $A \subseteq \Omega$, the probability P(A) belongs to some interval. Suppose Y_2 has no other information about the probability on Ω. In that case, Y_2 can only build a set $\mathcal{P}$ of probability measures on Ω compatible with the boundaries provided by Y_1. All that is known to Y_2 is that there exists a probability measure P and that $P \in \mathcal{P}$. Should Y_2 learn then that an event $A \subseteq \Omega$ has occurred, $\mathcal{P}$ should be updated to $\mathcal{P}_A$ where $\mathcal{P}_A$ is this set of conditional probability measures obtained by conditioning the probability measures $P \in \mathcal{P}$ on A (Smets, 1987, Fagin and Halpern, 1991a, Jaffray, 1992).

One obtains a similar result by assuming that one's belief is not described by a single probability measure as do the Bayesians but by a family of probability measures (usually the family is assumed to be convex). Conditioning on some event $A \subseteq \Omega$ is obtained as in the previous case.

7.1. Dempster's model.

A special case of upper and lower probabilities has been described by Dempster (1967, 1968). He assumes the existence of a probability measure on a space X and a one to many mapping M from X to Y. Then the lower probability of A in Y is equal to the probability of the largest subset of X such that its image under M is included in A. The upper probability of A in Y is the probability of the largest subset of X such that the images under M of all its elements have a non empty intersection with A. In the Artificial Intelligence community, this theory is often called the Dempster-Shafer theory.

7.2. Second order probabilities.

A generalization of an upper and lower probability theory to second-order probability theory is quite straightforward. Instead of just acknowledging that $P \in \mathcal{P}$, one can accept a probability measure P' on $\mathbb{P}_{\Omega}$, the set of probability measures on Ω. So for all $\mathcal{A} \subseteq \mathbb{P}_{\Omega}$, one can define the probability $P'(\mathcal{A})$ that the actual probability P on Ω belongs to the subset $\mathcal{A}$ of probability measures on Ω. In that case, the information $P \in \mathcal{P}$ induces a conditioning of P' into $P'(\mathcal{A}|\mathcal{P}) = P'(\mathcal{A} \cap \mathcal{P})/P'(\mathcal{P})$.

Second-order probabilities, i.e. probabilities over probabilities, do not enjoy the same support as subjective probabilities. Indeed, there seems to be no compelling reason to conceive a second-order probability in terms of betting and avoiding Dutch books. So the major justification for the subjective probability modeling is lost. Further introducing second-order probabilities directly leads to a proposal for third-order probabilities that quantifies our uncertainty about the value of the second-order probabilities.... Such iteration leads to an infinite regress of meta-probabilities that cannot be easily avoided.

7.3. Upper and lower probabilities and Dutch books.

The Dutch book argument requires that You are ready to take the position of the player or of the banker according to my decision. You are not allowed to ‘run away’. Bets are forced. Smith (1961) and Gilles (1982) have defended that bets might not be forced. In that case, You assess the maximal price You would be ready to pay in order to enter the game as the player, and the minimal price You would require from Your opponent in order You enter the same game as the banker. It seems perfectly reasonable that You would accept to pay at most .3$ to enter a game if You are the player and wins $1 if A occurs, whereas on the same time, You would not accept to be the Banker except if the opponent pays at least .8$ to enter the same game. In probability theory, the two prices should be equal to avoid a Dutch book, but now You can ‘run away’, so the Dutch books argument cannot be evoked. The maximal price You are ready to pay to enter the game as a player is the lower probability $P_*(A)$ You give to the event A. The minimal price You require from the opponent when You are the banker is the upper probability $P^*(A)$ You give to the event A. Your behavior is similar to the one You would have if all You knew is that the probability of A is in the interval $[P_*(A), P^*(A)]$ and You would be cautious. Cautious means that if You are the player, You assume that $P(A) = P_*(A)$, and if You are the banker, You assume $P(A) = P^*(A)$. In case the price would be fixed at some value between these two extremes, You would refuse both to be the player or the banker (and ‘run away’). Indeed for You the price would be too high to be the player, and too low to be the banker. Such a ‘run away’ behavior is forbidden in the classical Dutch book argument.

7.4. Other theories based on upper and lower probabilities:

Up to here, the theories underlying the use of upper and lower probabilities are based on the idea that uncertainty is represented by a unique probability function, but that You are unable to know exactly which probability function should be applied. There is some imprecision about the value of the probability function, but the probability function is assumed to exist somehow.

7.4.1. Families of probability functions.

Authors like Levi (1980) have defended that the state of belief of the agent cannot be described by a probability function, but a family of probability functions, without having to assume that one of the element of that family corresponds to some actual probability function. The concept of an ‘actual probability function’ is irrelevant. Belief states are more elaborate than suggested by the Bayesians. Such a theory fits nicely with the upper and lower probabilities described in the unforced bets contexts.

7.4.2. Inner and outer measures.

Fagin and Halpern (1991b), Voorbraak (1993) have studied the following problem. Suppose two algebras $\mathcal{A}$ and $\mathcal{B}$ defined on the set Ω, where $\mathcal{A}$ is a subalgebra of $\mathcal{B}$. Suppose the values of the probability measure are known only on the elements of the algebra $\mathcal{A}$. Fagin and Halpern try to determine the values of the probability measure on the subsets of the algebra $\mathcal{B}$. Because of the missing information, only the inner and outer measures for every event B in $\mathcal{B}$ can be determined. By construction, the inner (outer) measure is a lower (upper) probability function, and even a special one as the inner (outer) measure is a belief (plausibility) function, an obvious result when Ω is finite, and that is easily derived once Dempster’s one-to-many relation is considered.

8. Theories based on belief functions.

8.1. Credal versus pignistic levels.

The theories based on belief functions assume a major distinction between believing and acting. Uncertainty induces beliefs, i.e. graded dispositions that guide our behavior. These beliefs manifest themselves at two mental levels: the credal level where beliefs are entertained and the pignistic level where beliefs are used to act[3]. In probability theory, these two levels are not distinguished.

Authors like Ramsey (1931), Savage (1954), DeGroot (1970)... have indeed shown that if decisions must be "coherent", the uncertainty over the possible outcomes must be

[3] Credal and pignistic derive both from the latin words ‘credo’, I believe and ‘pignus’, a wage, a bet (Smith, 1961).

represented by a probability function. This result is accepted here, except that such *probability functions quantify the uncertainty only when a decision is really involved*. Uncertainty must be represented by a probability function at the pignistic level. This probability function is induced from the beliefs entertained at the credal level.

What we reject is the assumption that this probability function represents the uncertainty at the credal level. We assume that the pignistic and the credal levels are distinct. This implies that the justification for using probability functions at the credal level does not hold anymore (Dubois et al., 1996). In the transferable belief model (Smets, 1988), we defend that beliefs at the credal level are quantified by belief functions (Shafer, 1976). When decisions must be made, our belief held at the credal level induces a probability function at the 'pignistic' level. This probability function will be used in order to make decisions, using the expected utilities theory. The probability function is obtained by the so-called pignistic transformation the unique nature of which is derived and justified in Smets (1989). The probability function used at the pignistic level does not represent Your belief, it is a function induced by Your belief used only to compute the expected utilities.

8.2. The belief functions.

Maybe the easiest way to define a belief function is by starting with the so-called basic belief masses (bbm). The bbm m(A) given to the subset A of Ω is the amount of Your total unitary amount of belief that supports at most that the actual world is in A, and does not support any more specific subset of Ω because of a lack of information. So $m:2^{\Omega}\rightarrow[0,1]$ and

$$\sum_{A\subseteq\Omega} m(A) = 1$$

Suppose You receive the information that the actual world is in B. The specific support You gave to A is therefore transferred to A∩B. Indeed the reasons that lead You to believe at most that the actual world belongs to A are now reasons to believe that the actual world is in A∩B as You know that the actual world in indeed in B. This transfer of belief by the conditioning process explains the origin of the name we gave to the theory: the transferable belief model (TBM). This belief transfer is called the Dempster's rule of conditioning, except for a detail dealing with the outcome of the masses that where given to the subsets of $\overline{B}$, hence masses given to subsets that becomes ‘impossible’. In the TBM, we transfer them to the empty set. This mass represents the amount of internal conflict: indeed it is the amount of belief that was initially given to the subsets that turn out to be ‘impossible’. The conditioning information conflicts somehow with Your initial belief: You had some reasons to belief that the actual world belongs to a set $\overline{B}$ and now You learn that the actual world cannot belong to $\overline{B}$. The mass m(Ø) given to the empty set quantifies that conflict. A more subtle analysis of the nature of Ø and its mass m(Ø) is proposed in Perin (1997).

In Shafer's presentation, the bbm are proportionally normalized so that the empty set never gets a positive mass (Smets, 1992).

Suppose two subsets A and B of Ω where A is a subset of B. The bbm m(A) that supports that the actual world is in A also supports that the actual world is in B, indeed 'being in A' implies 'being in B'. Hence the amount of beliefs given to the fact that the actual world belongs to B is obtained by summing the bbm m(X) given to the subsets X of B. In that sum we do not included the empty set as m(Ø) would support B, but also $\overline{B}$. We define bel(B) as the total amount of *justified (necessary) specific support* given to B. It is obtained by summing all bbm given to subsets $A \in \mathfrak{R}$ with $A \subseteq B$ (and $A \neq Ø$).

$$\text{bel}(B) = \sum_{A: Ø \neq A \subseteq B} m(A)$$

We say *justified* (or necessary) because we include in bel(A) *only* the bbm given to subsets of A. For instance, consider two distinct atoms x and y of $\mathfrak{R}$. The bbm m({x,y}) given to {x,y} could support x if further information indicates this. However given the available information the bbm can only be given to {x,y}. We say *specific* because the bbm m(Ø) is not included in bel(A).

It can then be shown that the function bel so defined satisfied the following inequalities:

$$\forall n \geq 1,\ A_1, A_2, \ldots A_n \subseteq \Omega$$
$$\text{bel}(A_1 \cup A_2 \cup \ldots A_n) \geq \sum_i \text{bel}(A_i) - \sum_{i>j} \text{bel}(A_i \cap A_j) \ldots . - (-1)^n \text{bel}(A_1 \cap A_2 \cap \ldots A_n) \qquad (8.1)$$

Related to bel, one can also define the so-called plausibility function pl where pl(A) quantifies the maximum amount of *potential specific support* that could be given to the fact that the actual world belongs to A. It is obtained by adding all those bbm given to subsets A compatible with B, i.e., such that $A \cap B \neq Ø$:

$$\text{pl}(B) = \sum_{A:\ A \cap B \neq Ø} m(A) = \text{bel}(\Omega) - \text{bel}(\overline{B})$$

We say *potential* because the bbm included in pl(A) could be transferred to some non-empty subsets of A if new information could justify such a transfer.

The bel and pl functions are necessary and potential support functions. Full details on the TBM is given in (Smets, 1988, 1994, Smets and Kennes, 1994).

The literature dealing with belief functions is poised by a serious confusion that often leads to erroneous results. In the TBM, the values of bel do not result from some probability. The theory for quantifying the strength of Your belief that the actual world belongs to the subsets of Ω is developed and justified without considering the existence of some underlying, maybe hidden, probability. In Shafer's book (1976), the same approach prevails. But in the early 80's, authors understood the approaches based on belief functions as a theory of upper and lower probability. Indeed it is mathematically true that given a normalized belief function bel on Ω, it is always possible to define a family *P* of probability functions P defined on Ω that satisfy the following constraints:

$$\forall P \in \mathcal{P}, \forall A \subseteq \Omega, \qquad bel(A) \leq P(A) \leq pl(A)$$

This property has often been used to claim that belief functions are just lower probability functions. The danger of that idea is that one extends the statement by claiming that belief functions concerns an ill known probability function, in which case one assumes the existence of a probability function P that belongs to *P* and that represents 'something', and the 'something' is of course understood as Your degree of belief. At the static level, the difference is the following. In the TBM, bel represents Your beliefs. In the lower probability approach, one assumes that Your belief is represented by a probability function, which value is only known to belong to *P*, and bel is just the lower limit of that family *P*.

The difference becomes more obvious once conditioning on an event X is introduced. In the TBM, conditioning of bel on X into bel_X is achieved by Dempster's rule of conditioning, hence by the transfer of the bbm as explained above. In the lower probability approach, the conditioning is obtained by considering every probability function P in *P*, conditioning P on X and collecting them in a new family $\mathcal{P}_X$ of conditional probability functions P_X. The results are different in that bel_X is not the lower envelope of $\mathcal{P}_X$ (Kyburg, 1987, Voorbraak, 1991).

The family *P* of probability functions compatible with a given belief function has nevertheless a meaning in the TBM, but quite different from the one considered in the lower probability approach. Given a belief function, the probability function used to compute the expected utilities at the pignistic level when a decision is involved is computed by the so called pignistic transformation. The result depends of course on bel, but also on the betting frame, i.e., the elementary options considered in the decision process. Suppose we let the betting frame varies. For each possible betting frame we get a probability function. Collect all these probability functions into a family. This family is the same as the family *P* (Wilson, 1993). So we can derive *P* in the TBM. The difference with the lower probability approach is that we start with bel and derive *P* as a by product, whereas the lower probability approach starts with *P* and derive bel as a by product.

8.3. Probabilities defined on modal propositions.

Classically probability theory is defined on propositional logic. The whole presentation of probability theory could be realized by using propositions instead of events and subsets. So for a proposition p, P(p) would be the probability that p is true (hence that p is true in the actual world). Extending the domain of the probability functions to modal propositions is quite feasible. Ruspini (1986) examines the ‘probability of knowing’. Pearl (1988) examines the ‘probability of provability’. Both approaches fit essentially with the same ideas.

The probability $P(\Box p)$ is the probability that $\Box p$ is true in the actual world. The worlds of Ω can be partitioned in three categories: those where $\Box p$ holds, those where $\Box\neg p$ holds, and those where neither $\Box p$ nor $\Box\neg p$ hold. Hence,

$$P(\Box p) + P(\Box\neg p) + P(\neg\Box p \& \neg\Box\neg p) = 1. \tag{8.2}$$

Suppose You define bel(p) as $P(\Box p)$, i.e., You define bel(p) as the probability that p is proved, is known, is necessary, depending on the meaning given to the $\Box$ operator. The equality (8.2) becomes then:

$$\mathrm{bel}(p) + \mathrm{bel}(\neg p) \leq 1,$$

Similarly the other inequalities described with belief functions (8.1) are also satisfied. This approach provides a nice interpretation of bel as the probability of provability, of knowing, etc... Nevertheless the theory so derived is not the TBM, as seen once conditioning is involved (Smets, 1991a). The probability $P(\Box p|\Box q)$ of knowing p when knowing q is:

$$P(\Box p|\Box q) = \frac{P(\Box p \& \Box q)}{P(\Box q)} = \frac{P(\Box(p\&q))}{P(\Box q)} = \frac{\mathrm{bel}(p\&q)}{\mathrm{bel}(q)}$$

This is not Dempster's rule of conditioning. It happens to be the geometrical rule of conditioning (Shafer, 1976b). Dempster's rule of conditioning is obtained if the impact of the conditioning event results in an adaptation of the accessibility relation underlying the modal logic (Smets, 1991). Conceptually, probabilities defined on modal propositions belongs to the family of theories initially studied by Dempster (1967) where a probability measure and a one-to-many mapping are assumed. It provides a nice framework and explains the origin of the conditioning rules.

8.4. Dempster-Shafer theories.

In the literature, one often encounters the phrase ‘Dempster-Shafer theory’ (Gordon and Shortliffe, 1984). Unfortunately what is covered by such a label is often quite confusing, and varies widely from authors to authors (Smets, 1994). Most often, it corresponds to the lower probability approach described above. Sometimes it corresponds to Dempster’s original approach (Shafer, 1992). The hints theory of Kohlas (Kohlas, 1993) (see section 8.5) corresponds to that last definition. Dempster’s theory can also be seen as a random set model, i.e., a theory based on a one-to-many mapping between a probability space and another space. The concept of a random variable we would have if the mapping was one-to-one, is generalized into the concept of a random set.

8.5. The theory of hints.

Kohlas and Monney (1995) have proposed a theory of hints. They assume Dempster's original structure (Ω, P, Γ, Θ) where Ω and Θ are two sets, P is a probability measure on Ω and Γ is a one-to-many mapping from Ω to Θ. They assume a question, whose answer is unknown. The set Θ is the set of possible answers to the question. One and only one element of Θ is the correct answer to the question. 'The goal is to make assertions about the answer in the light of the available information. We assume that this information allows for several different interpretations, depending on some unknown circumstances. These interpretations are regrouped into the set Ω and there is exactly one correct interpretation. Not all interpretations are equally likely and the known probability measure P on Ω reflects our information in that respect. Furthermore, if the interpretation $\omega \in \Omega$ is the correct one, then the answer is known to be in the subset $\Gamma(\omega) \subseteq \Theta$. Such a structure $\mathcal{H} = (\Omega, P, \Gamma, \Theta)$ is called a hint... An interpretation $\omega \in \Omega$ supports the hypothesis H if $\Gamma(\omega) \subseteq H$ because in that case the answer is necessarily in H. The degree of support of H, denoted sp(H), is defined as the probability of all supporting interpretation of H' (Kohlas and Monney, 1995, page vi).

The hints theory corresponds to Dempster's original approach. They call their measure a degree of support, instead of belief, to avoid personal, subjective connotation, but degrees of support and degrees of belief are mathematically equivalent and conceptually very close. In the hints theory, the primitive concept is the hint from which degrees of supports are deduced, whereas the TBM and Shafer's initial approach (Shafer, 1976), consider the degrees of belief as a primitive concept. The hints theory is quite similar to the probability of provability theory (section 8.3). All details on the hints theory can be found in Kohlas and Monney (1995).

8.6. Axiomatisation of the TBM.

Several axiomatic justifications for representing quantified beliefs by belief functions have been suggested. Initially, Shafer (1976a) assumes that any measure of belief bel satisfy the inequalities:

$\forall n \geq 1,\ A_1, A_2, \ldots A_n \subseteq \Omega$

$$\mathrm{bel}(A_1 \cup A_2 \cup \ldots A_n) \geq \sum_i \mathrm{bel}(A_i) - \sum_{i>j} \mathrm{bel}(A_i \cap A_j) \ldots . - (-1)^n \mathrm{bel}(A_1 \cap A_2 \cap \ldots A_n)$$

In the transferable belief model (Smets and Kennes, 1994), we assume the existence of parts of beliefs (the bbm) that support a proposition without supporting any more specific proposition. Both approaches are strictly equivalent. We introduce the second in response to the criticism that the inequalities of Shafer were too artificial and difficult to accept as a natural requirement for a measure of belief, and hope ours are more ‘palatable’.

Wong et al. (1990) have presented an axiomatic framework based on the representation of a belief-order relation ≥ (>) where B≥C (B>C) means ‘B is not less believed than C’ (‘B is more believed than C’). They replace the disjoint union axiom assumed in probability theory (Koopman, 1940, Fine, 1973):

$$A\cap(B\cup C)=\emptyset \Rightarrow (B\geq C \Leftrightarrow A\cup B\geq A\cup C)$$

by a less restrictive axiom:

$$C\subseteq B,\ A\cap B=\emptyset \Rightarrow (B>C \Rightarrow A\cup B\geq A\cup C)$$

Under this last axiom, the ≥ belief-ordering can always be represented by a belief function. Unfortunately, other functions like the monotone capacities of order 2 (Choquet, 1953) can also represent the ≥ ordering.

In Smets (1993a, 1997) we propose a full axiomatization for the representation of quantified belief. We present ‘natural requirements’ that justify the behavior of the belief functions under conditioning, refinement and coarsening. It is also easy to show that any function used to represent quantified beliefs must be a monotone capacity of order 2. Several extra requirements are proposed to show that they must be monotone capacities of infinite order (i.e., belief functions), but they are not as compelling as the other requirements.

8.7. Degenerating the theories on a {0,1} range.

Is it often enlightening to examine what become the theories we have studied when their range is reduced to the set {0, 1} instead of the whole interval [0, 1]. In probability theory, for every proposition p we have only two states:
P(p) = 1 : p is believed.
P(p) = 0 : p is not believed, i.e., ¬p is believed.
Ignorance cannot be represented.

In possibility theory, we have three states:
N(p) = 1 and ∏(p) = 1 : p is necessary
N(p) = 0 and ∏(p) = 1 : p is contingent (and so is ¬p).
N(p) = 0 and ∏(p) = 0 : p is impossible (hence ¬p is contingent or necessary).
With belief functions, the same three states are achieved.

In fact probability theory is a generalization of a highly degenerated modal logic where $\Box = \Diamond$, whereas possibility theory and belief functions theory generalized classical modal logic and only consider that $\Box p \supset \Diamond p$ holds (Mongin, 1992)

9. Conclusions: a word of ecumenism.

After this survey of the theories proposed to quantify uncertainly, it seems interesting to make a break for asking what to do in practice. How can we select the adequate approach? The matter is hardly settled and no general solution has yet been advanced. Usually theories are chosen for either their efficiency, which theory is easy to implement, either for their behavior, which one works nicely, and sometimes for their adequacy, which one fits with the problem at hand. Efficiency should lead to possibility theory as all it needs is a fast min-max processor. Good behavior is hardly an adequate arguments, it allows to reject inconsistent theories, but does not help in choosing one as 'good' is hardly defined. Finally adequacy is too often completely neglected. It would require a close examination of the axioms underlying the various theories and an assessment of their adequacy to represent the problem at hand.

The major error is in selecting automatically or blindly an approach without even considering that it might be inadequate. This happens with probability theory often chosen just by tradition, because everybody does so, or even by sheer ignorance of the other methods. We plead for open-minded attitude, hoping this paper and those in these handbooks will provide tools for choosing in a more rational way which theory is to be used for which problems.

Bibliography.

ADAMS E.W. (1975) The logic of conditionals. Reidel, Dordrecht.

BENFERHAT S., DUBOIS D. and PRADE H. (1992) Representing default rules in possibilistic logic. Proc. 3rd Conf. Prionciple of Knowledge Representation and Reasoning, KR'92.

BENFERHAT S., SAFFIOTTI A. and SMETS Ph. (1995) Belief functions and default reasoning. in Uncertainty in AI 95. Besnard Ph. and Hanks S. eds. Moragan Kaufman, San Francisco, Ca, pg. 19-26.

BESNARD P., JAOUEN P. and PERIN J.Ph. (1996) Extending the transferable belief model for inconsistency handling. IPMU-96, 143-148.

BRADLEY R. and SWARTZ N. (1979) Possible worlds. Basil Blackwell, Oxford, UK.

CARNAP R. (1950) Logical Foundations of Probability. University of Chicago Press, Chicago, Illinois.

CARNAP R. (1952) The continuum of inductive methods. University of Chicago Press, Chicago, Illinois.

CHELLAS B.F. (1980) Modal logic. Cambridge Univ. Press, G.B.

CHOQUET G. (1953) Theory of capacities. Annales de l'Institut Fourier, Université de Grenoble, 5:131-296.

COX R.T. (1946) Probability, frequency and reasonable expectation. Amer.J.Phys. 14:1-13.

CLARKE M.R.B., FROIDEVAUX C. GREGOIRE E. and SMETS Ph. (1991) Guest Editors of the Special Issue on 'Uncertainty, Conditional and Non Monotonicity. Positions and Debates in Non-Standard Logics.'. J. Applied Non-Classical Logics. Vol.1, N° 2.pg. 103-310.
DE FINETTI B. (1974) Theory of Probability. Wiley, London. Vol.1 and Vol. 2,
DEGROOT M.H. (1970) Optimal statistical decisions. McGraw-Hill, New York.
DEMPSTER A.P. (1967) Upper and lower probabilities induced by a multplevalued mapping. Ann. Math. Statistics 38:325-339.
DEMPSTER A.P. (1968) A generalization of Bayesian inference. J. Roy. Statist. Soc. B.30:205-247.DUBOIS D., GARBOLINO P., KYBURG H.E., PRADE H. and SMETS Ph. (1991) Quantified Uncertainty. J. Applied Non-Classical Logics 1:105-197.
DUBOIS D. and PRADE H. (1980) Fuzzy sets and systems, theory and applications. Academic Press.
DUBOIS D. and PRADE H. (1982) A class of fuzzy measures based on triangular norms. Int. J. General Systems 8:43-61.
DUBOIS D. and PRADE H. (1985) A review of fuzzy sets aggregation connectives. Information Sciences 36:85-121.
DUBOIS D. and PRADE H. (1988) Possibility theory. Plenum, London.
DUBOIS D. and PRADE H. (1992) Evidence, knowledge and belief functions. Int. J. Approx. Reasoning 6:295-320.
DUBOIS D. and PRADE H. (1991) Epistemic entrenchment and possibilistic logic. Artig. Intell. J. 50:223-239.
DUBOIS D., MORAL S. and PRADE H. (1995) A semantics for possibility theory based on likelihoods. Proc. FuzzIEEE95, Yokohama. (under press in J. Math. Anal. Appl.)
DUBOIS D., PRADE H. and SMETS Ph. (1996) Representing partial ignorance. IEEE System Machine and Cybernetic.
DUBOIS D. and PRADE H. (1997) Possibility theory and possibilistic logic: an introduction. This handbook, vol. 1.
EARMAN J. (1992) Bayes or Bust?. A Critical Examination of Bayesian Confirmation Theory. MIT Press, Cambridge, Mass.
EDWARDS A.W.F. (1972) Likelihood. Cambridge University Press, Cambridge, UK
FAGIN R. and HALPERN J. (1991a) A new approach to updating beliefs. in Uncertainty in Artificial Intelligence 6 (Bonissone P.P., Henrion M., Kanal L.N. and Lemmer J.F. eds.) North Holland, Amsteram, 347-374.
FAGIN R. and HALPERN J. (1991b) Uncertainty, belief and probability. Computational Intell. 7:160-173.
FETZER J. (1971) Dispositional probabilities. Boston Studies in the Philosophy of Science 8:473-482.
FAGIN R., HALPERN J., MOSES Y. and VARDI M.Y. (1995) Reasoning about knowledge. MIT Press Cambridge, Mass.
FINE T. (1973) Theories of probability. Academoic Press, New York.
GABBAY D. and HUNTER A. (1991) Making inconsistency respectable 1: A logical framework for inconsistency in reasoning. In Ph Jorrand and J Kelemen, editors, Fundamentals of Artificial Intelligence, volume 535 of LNCS, pp 19-32, Springer.
GÄRDENFORS P., HANSSON B. and SAHLIN N.E. (1983) Evidentiary value: philosophical, judicial and psychological aspects of a theory. C.W.K. Gleerups, Lund, Sweden.
GÄRDENFORS P. (1988) Knowledge in flux. Modelling the dynamics of epistemùic states. MIT Press, Cambridge, Mass.
GEBHARDT J. and KRUSE R. (1993) The Context Model : An Integrating View of Vagueness and Uncertainty, Int. J. of Approximate Reasoning 9, 283-314.
GIGERENZER G. (1996) Psychological challenge for normative models. Manuscript.
GILES R. (1982) Foundation for a possibility theory. in Fuzzy Information and Decision Processes, GUPTA M.M. and SANCHEZ E. eds. North Holland, Amsterdam, 183-195.
GOOD I.J. (1950) Probability and the weighting of evidence. Hafner.

GOOD I.J. (1983) Good thinking: the foundations of probability and its applications. Univ. Minnesota Press, Minneapolis.
GOODMAN I.R., NGUYEN H.T. and WALKER E.A. (1991) Conditional inference and logic for intelligent systems. Elsevier, Amsterdam.
GORDON J. and SHORTLIFFE E.H. (1984) The Dempster-Shafer theory of evidence, in BUCHANAN B.G. and SHORTLIFFE E.H. (Eds), Rule-Based Expert Systems: the MYCIN experiments of the Stanford Heuristic Programming Project. Addison-Wesley, Reading, MA. 272-292.
HACKING I. (1965) Logic of statistical inference. Cambridge University Press, Cambridge, U.K.
HACKING I.(1975) The emergence of probability. Cambridge University Press, Cambridge.
JAFFRAY J.Y. (1992) Bayesian updating and belief functions. IEEE Trans. SMC, 22:1144-1152.
JAOUEN P. (1997) Modélisation de contradictions par extension de la théorie de l'évidence. Thesis, Université de Rennes 1.
JEFFREY R.C. (1983) The logic of decision. 2nd Ed. Univ. Chicago Press, Chicago.
JEFFREY R. (1988) Conditioning, kinematics, and exchangeability. in Causation, Chance, and Credence, Skyrms B. and Harper W.L. eds., Reidel, Dordrecht, vol.1, 221-255.
KEYNES J.M. (1921) A Treatise on Probability. Macmillan, London.
KOHLAS J. (1993) A Mathematical Theory oh Hints. Int. J. Gen. Syst.
KOHLAS J. and MONNEY P.A. (1995) A Mathematical Theory of Hints. An Approach to Dempster-Shafer Theory of Evidence. Lecture Notes in Economics and Mathematical Systems No. 425. Springer-Verlag.
KOOPMAN B.O. (1940) The bases of probability. Bull. Amer. Math. Soc. 46:763-774.
KYBURG H.E.Jr. (1987) Bayesian and non-Bayesian evidential updating. Artificial Intelligence, 31:271-294.
LEA SOMBE (1989) Raisonnements sur des informations incomplètes in intelligence artificielle. Teknea, Toulouse. English translation: Reasoning under incomplete information in artificial intelligence. Wiley, New York (1990).
LEA SOMBE (1994) A glance at revision and updating in knowledge bases. Inter. J. Intelligent Systems, 9:1-28.
LEVI I.(1980) The enterprise of knowledge. MIT Press, Cambridge, Mass.
MELLOR D.H. (1969) Chance. Proc. Aristotelian Soc. Suppl. vol. page 26.
MONGIN Ph. (1992) Some connections between epistemic logic and the theory of nonadditive probability. Personal communication.
PARIS J. (1994) The uncertainty reasoner’s companion. Cambridge Univ. Press., Cambridge, GB.
PEARL J. (1988) Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann Pub. San Mateo, Ca, USA.
PEARL J. (1990a) System Z: a natural ordering of defaults with tractable applications to default reasoning. Proc. TARK'90, 121-135.
PEARL J. (1990b) Reasoning with Belief Functions: an Analysis of Compatibility. Intern. J. Approx. Reasoning, 4:363-390.
PERIN J.Ph. (1997) Théorie de l’évidence: modélisation et application pour un SIC. Thesis, Université de Rennes 1.
POLLOCK J.L. (1990) Nomic probability and the foundations of induction. Oxford University Press, New York.
POPPER K.R. (1959a) The Logic of Scientific Discovery. New York, Basic Books.
POPPER K. (1959b) The propensity interpretation of probability. British J. Phil. Sc: 10:25-42.
RAMSEY F.P. (1931) Truth and probability. in Studies in subjective probability, eds. KYBURG H.E. and SMOKLER H.E., p. 61-92. Wiley, New York.
REICHENBACH H. (1949) The Theory of Probability. Berkeley: University of California Press.
RUSPINI E.H. (1986) The logical foundations of evidential reasoning. Technical note 408, SRI International, Menlo Park, Ca.

SAVAGE L.J. (1954) Foundations of Statistics. Wiley, New York.
SCHUM D.A. (1994) Evidential foundations of probabilistic reasoning. Wiley, New York.
SHACKLE G.L.S. (1969) Decision, order, and time in human affairs. Cambridge Univ. Press, Cambridge.
SHAFER G. (1976) A mathematical theory of evidence. Princeton Univ. Press. Princeton, NJ.
SHAFER G. (1976b) A theory of statistical evidence. in Foundations of probability theory, statistical inference, and statistical theories of science. Harper and Hooker ed. Reidel, Doordrecht-Holland.
SHAFER G. (1992) Rejoinder to Comments on "Perspectives in the theory and practice of belief functions". Intern. J. Approx. Reasoning, 6:445-480.
SMETS Ph. (1981) The Degree of Belief in a Fuzzy Event. Information Sciences 25 : 1-19.
SMETS Ph. (1982a) Probability of a Fuzzy Event : an Axiomatic Approach. Int. J. Fuzzy Sets and systems, 7 : 153-164.
SMETS Ph. (1982a) Possibilistic Inference from Statistical Data. In : Second World Conference on Mathematics at the Service of Man. A. Ballester, D. Cardus and E. Trillas eds. Universidad Politecnica de Las Palmas, pp 611-613.
SMETS Ph. (1987) Belief functions and generalized Bayes theorem. Second IFSA Congress, Tokyo, Japan, July 20-25, 1987. pg. 404-407.
SMETS Ph. (1988) Belief functions. in Non Standard Logics for Automated Reasoning, ed. Smets Ph., Mamdani A., Dubois D. and Prade H. Academic Press, London pg253-286.
SMETS Ph. (1989) Constructing the pignistic probability function in a context of uncertainty. Uncertainty in Artificial Intelligence 5, Henrion M., Shachter R.D., Kanal L.N. and Lemmer J.F. eds, North Holland, Amsterdam, 29-40
SMETS Ph. (1991a) Probability of provability and belief functions. Logique et Analyse, 133-134:177-195.
SMETS Ph. (1991b) Implication and modus ponens in fuzzy logic. in GOODMAN I.R., GUPTA M.M., NGUYEN H.T. and ROGERS G.S. Conditional logic in expert systems. Elsevier, Amsterdam, pg. 235-268.
SMETS Ph. (1992) The nature of the unnormalized beliefs encountered in the transferable belief model. in Dubois D., Wellman M.P., d'Ambrosio B. and Smets Ph. Uncertainty in AI 92. Morgan Kaufmann, San Mateo, Ca, USA, 1992, pg.292-297.
SMETS Ph. (1993a) An axiomatic justifiaction for the use of belief function to quantify beliefs. IJCAI'93 (Inter. Joint Conf. on AI), Chambery. pg. 598-603.
SMETS Ph. (1993b) Probability of Deductibility and Belief Functions. ECSQARU 93.
SMETS Ph. (1993c) No Dutch Book can be built against the TBM even though update is not obtained by Bayes rule of conditioning. SIS, Workshop on Probabilisitic Expert Systems, (ed. R. Scozzafava), Roma, pg. 181-204..
SMETS Ph. (1993d) Jeffrey's rule of conditioning generalized to belief functions. Uncertainty in AI 93. Heckerman D. and Mamdani A. eds. Morgan Kaufmann, San Mateo, Ca, USA, 500-505.
SMETS Ph. (1994) What is Dempster-Shafer's model? in Advances in the Dempster-Shafer Theory of Evidence. Yager R.R., Kacprzyk J. and Fedrizzi M., eds, Wiley, New York, pg. 5-34..
SMETS Ph. (1997) The axiomatic justification of the transferable belief model. Artificial Intelligence 92: 229-242.
SMETS Ph. and KENNES R. (1994) The transferable belief model. Artificial Intelligence 66:191-234.
SMETS Ph., MAGREZ P. (1987) Implication in Fuzzy Logic. Int. J. Approximate Reasoning 1:327-348.
SMETS Ph., MAGREZ P. (1988) The Measure of the Degree of Truth and of the Grade of Membership. Int. J. Fuzzy Sets and Systems. 25:67-72.
SMITH C.A.B. (1961) Consistency in statistical inference and decision. J. Roy. Statist. Soc. B23:1-37.

SMITH C.A.B. (1965) Personal probability and statistical analysis. J. Roy. Statist. Soc. A128, 469-499.
SPOHN W. (1990) A general non-probabilistic theory of inductive reasoning. in Uncertainty in Art. Intel. 4: Schacter R.D., Levitt T.S., Kanal L.N. and Lemmer J.F. (eds). North Holland, Amsterdam, pg. 149-158.
SUGENO M. (1977) Fuzzy measures and fuzzy integrals: a survey. in M.M. Gupta, G.N. Saridis and B. R. Gaines, eds., Fuzzy Automata and Decision Processes. North Holland, Amsterdam.89-102.
TELLER P. (1973) Conditionalization and Observation. Synthesis 26:218-258.
TELLER P. (1976) Conditionalization, Observation and Change of Prefetrence. in W Harper and C.A. Hooker, eds., Foundations of Probability Theory, Statistical Inference, and Statistical Theory of Science. Reidel, Doordrecht. 205-259.
THOMAS S.F. (1979) A theory of semantics and possiblle inference with applications to decisions analysis. Ph. D. Thesis, Univ. Toronto.
THOMAS S.F. (1995) Fuzziness and probability. ACG Press, Wichita, KS, USA.
VON MISES R. (1957) Probablity, Statistics and Truth. (Second Ed.) London, Allen and Unwin.
VOORBRACHT F. (1991) On the justification of Dempster's rule of combination. Artificial Intelligence 48:171-197.
VOORBRAAK F. (1993) As Far as I Know: Epistemic Logic and Uncerttaiunty. Dissertation, Utrecht University.
WALLEY P. (1991) Statistical reasoning with imprecise probabilities. Chapman and Hall, London.
WEBER S. (1984) Decomposable measures and integrals for Archimdian t-conorms. J. Math. Anal. Appl. 101:114-138.
WILSON N. (1993) Decision making with belief functions and pignistic probabilities. in Symbolic and Quantitative Approaches to Reasoning and Uncertainty. Clarke M., Kruse R. and Moral S. (eds.), Springer Verlag, Berlin, pg. 364-371.
WONG S.K.M., YAO Y.Y., BOLLMANN P. and BÜRGER H.C. (1990) Axiomatization of qualitative belief structure. IEEE Trans. SMC, 21:726-734.
YAGER R. (1991) Connectives and quantfiers in fuzzy sets. Int. J. Fuzzy Sets and Systems 40:39-76.
ZADEH L.A. (1965) Fuzzy sets. Inform.Control. 8:338-353.
ZADEH L. (1968) Probability measures of fuzzy events. J. Math. Anal. Appl. 23: 421-427.
ZADEH L.A. (1973) Outline of a new approach to the analysis of complex systems and decision processes. IEEE SMC 3:28-44.
ZADEH L. (1978) Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems 1: 3-28.