Non Standard Probabilistic and Non Probabilistic Representations of Uncertainty.

Philippe SMETS[1]
I.R.I.D.I.A. Université Libre de Bruxelles
50 av. Roosevelt, CP 194-6. B-1050, Brussels, Belgium.

Summary: Survey of the mathematical models proposed to represent quantified beliefs, and their comparison. The models considered are separated into non standard probability and non probability models, according to the fact they are based on probability theory or not. The first group concerns the upper and lower probability models, the second the possibility theory and the transferable belief model.

Keywords: uncertainty, belief representation, probability functions, upper and lower probability functions, possibility functions, belief functions.

1. Introduction.

This paper surveys the mathematical models that have been proposed to represent quantified beliefs. The mathematical representation of the ‘real world’ has been a permanent subject of research as it provides an objective and unambiguous formalization of the ‘reality’. That they represent idealized reality does not reduce their value. The 19th. century has seen the development of mathematical models to represent physical realities. In the 20th. century, randomness became omnipresent. More recently, the modelization has been extended toward the representation of subjectivity[2], and in particular of uncertainty.

Uncertainty can take numerous forms (Smets, 1991a, 1993a, Smithson, 1989) and usually induces ‘beliefs’, i.e. the graded dispositions that guide ‘our’ behavior. Proposed models of belief can be split into symbolic models as in modal logic (Hintikka, 1962, Chellas, 1980) or numerical models as those studied here. The model we study are normative, the

---

[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.
[2] Nguyen H., 1994, personal communication.

agent that holds the beliefs, called You hereafter, is an ‘ideal rational subject’. That humans hardly behave as predicted by the normative model reflects essentially their lack of ‘rationality’, a well known fact that does not bear negative connotation. The ‘rationality’ that underlies the models and the so called human rationality do not cover the same concepts.

The Bayesian model based on probability functions was the first model proposed to represent quantified beliefs. In spite of its success, alternative models have been proposed. They can be split into two families: the non-standard probability and the non-probability models (Kohlas, 1994b). The models based on non-standard probabilities are extensions of models based on probability functions. They include the upper and lower probabilities models (Good, 1950, Smith, 1961, Kyburg, 1987, Walley, 1991, Voorbraak, 1993), Dempster-Shafer's models (Dempster, 1967, Shafer, 1976a, Smets, 1994b), the Hints models (Kohlas and Monney, 1994a), the probability of provability models (Ruspini, 1986, Pearl, 1988, Smets, 1991c),. The non-probabilistic models are not based on probability functions, but on alternative functions like the possibility functions and the belief functions. They include the transferable belief model (Smets, 1988, 1990a, Smets and Kennes, 1994), the possibility theory model (Zadeh, 1978, Dubois and Prade, 1985). The difference between these models and their domain of application is discussed hereafter.

We first define what concept of belief is covered. Belief represent the opinion of the agent, You, at time t given what You know. We can say “You believe at level .7 that the name of Philippe Smets’ daughter is Karine”. It means that “You believe at level .7 that the proposition ‘the name of Philippe Smets’ daughter is Karine’ is true”. So beliefs weight the strength given by the agent to the fact that a given proposition is true. The domain of beliefs is the truth status of a proposition. When the truth domain is restricted to True and False, shortcuts can be used without danger, and we can just say “You believe p at .7” where p is a proposition. When the truth domain becomes more exotic, like in multi-valued logics and in fuzzy logics, these shortcuts can be misleading, and the use of the full expressions will avoided ambiguities, if not plain stupidities.

Coming back to classical logic, the meaning of the sentence ‘You believe at level .7 that it will rain tomorrow’ can be either: ‘the measure of the belief held by You that the proposition "it will rain tomorrow" is true is .7’ or: ‘You believe at level .7 that “tomorrow” belongs to the set of rainy days’. So beliefs given to ‘propositions’ can equivalently be given to the subset of worlds that denote the propositions. Defining beliefs on propositions or on sets of possible worlds is equivalent. We will restrict ourselves to propositions and possible worlds, and left aside more elaborate structures.

What are the needed ingredients for a model for the representation of quantified beliefs? We must be able to represent static states of belief. Such static states representation should be able to cover every states of belief, from the state of total ignorance up to the state of absolute knowledge. We must describe the dynamic of beliefs, i.e., how a new piece of information changes the belief states. Decisions are the only observable outcomes of a belief state and the most important practical components of any model for uncertainty. So we must also explain how decisions will be made. We will see how the various models answer to these questions.

In Section 2, we present and criticize the justifications that lead to the traditional Bayesian model based on probability functions. In section 3, we compare various non standard probability models. In section 4, we present the major non probability models: the possibility theory and the transferable belief model. In section 5, we just conclude with small remarks.

2. The Probability Models.

The representation of quantified beliefs really started in the eightieth century with the work of Bernoulli, De Moivre and Lambert (Shafer, 1978) culminating in the famous essay of Bayes (1763) where the author defends that beliefs should be mathematically represented by probability functions. Today three types of justifications are usually proposed in defense for such a representation: the decision-oriented, the preference-ordering and Cox algebraic justifications.

2.1. Decision-oriented justifications.

Ramsey (1931) defines a probability as the price one would be ready to pay to bet on an event which reward is $1 if the event occurs and nothing if the event does not occur. Such a definition is usually based on a betting scheme in which there is a player and a banker. The player must pay to the banker a certain amount of money (say $p) to enter a game where the banker will pay $1 to the player if event A occurs and $0 if A does not occur. The probability assessor (You) must decide the price $p of the game and I (Your opponent) decide which position will be held by You, banker or player. Besides, the bet is "forced" in that You must play. The prices You give to each game are arbitrary except that they must be coherent, i.e. they must satisfy the axioms of probability theory. If they did not, then I could always build a set of games where You would loose for sure.

For instance, if You decide to pay $.6 on a game where the winning event is A (Game 1) and $.3 on a game where the winning event is $\overline{A}$ (Game 2), then I would force You to be the banker for both games. If A occurs You loose \$.4 on Game 1 (You receive \$.6 and pay \$1) and You win \$.3 on Game 2 (You receive \$.3 and pay \$0), hence Your net loss is \$.1. If $\overline{A}$ occurs You win \$.6 on Game 1 and You loose \$.7 on Game 2, hence Your net loss is also \$.1. Therefore in both cases, You suffer a loss.

This strategy that leads to a sure loss is called a Dutch Book strategy. In order to avoid such a Dutch Book, it has been shown that the prices given to the various games must satisfy the axioms of probability theory, in particular, the additivity and the conditioning rules. In order to derive the conditioning rule, the set of potential events are partitioned in 3 subsets: A, B, C. Bets are posted such that You win if A occurs, You loose if B occurs and the game is canceled if C occurs in which case the banker returns the ticket price to the player. Such bet is called a conditional bet and leads to the definition of the conditional probability of A given $A \cup B$: $P(A|A \cup B) = P(A)/P(A \cup B)$

The decision-oriented approach seems quite convincing, except for the fact that it is based on "forced" bets. Some critics were raised against it as people feel they are not forced to bet. It leads to a generalization of the probability models where the concepts of upper and lower probabilities are introduced (Walley, 1991). The lower probability of an event is defined as the maximum amount You would be ready to pay to enter the game where You win \$1 if the event occurs and nothing otherwise. It does not require that You would accept to be the banker. The upper probability is defined as the minimum amount You would require that the player pays to enter the game if you were the banker. So when the probability assessor can assign prices without committing himself to accept to be the player or the banker according to his opponent's orders, the Dutch Book argument collapses (Smith, 1961, Williams, 1976, Jaffray, 1989, Voorbraak, 1993).

Another critic is oriented towards the idea that what Dutch Book argument leads to are not a quantified representation of our belief but a quantified representation of our belief qua basis of decision (Ramsey, 1931). Beliefs held outside any decision context are not considered in the Dutch Book argument and therefore the argument does not justify that quantified belief in general must be represented by a probability function. It only says that when decision must be made, beliefs will induce a probability measure that will be used for the decision-maker. Therefore, the question is to decide if beliefs can exist outside any decision process.

2.2. Credal versus pignistic beliefs.

When developing the transferable belief model (see section 4.2), we introduced a two-level mental model for representing belief. At one level beliefs are entertained without any regard to decision-making and at the other level beliefs are used to make decision. We call these two mental levels, the credal level and the pignistic level (respectively from credo = I believe and pignus = a bet, both in Latin, (Smith, 1961)). That beliefs are necessary ingredients for our decisions does not mean that beliefs cannot be entertained without any revealing behavior manifestations (Smith and Jones, 1986, p.147). A belief is a disposition to feel that things are thus-and-so. It must be contrasted with the concept of acceptance (Cohen, 1993). A probability measure is a tool for action, not for assessing strength of evidence (Sahlin, 1993).

Some authors will just discard the existence of a credal level arguing that only the pignistic level exists and indeed, at that level, beliefs should better be represented by a probability measure if one wants to avoid Dutch Books or similar sub-optimal behaviors. We will defend the distinction between the two levels and the whole paper is devoted to the development of a quantified representation of belief at the credal level. We will illustrate the impact of considering the two-level model in the transferable belief model through the 'Peter, Paul and Mary Saga'.

2.3. The preference-ordering justification.

Arguments based on preference-ordering have been proposed in order to support the use of probability measures for quantified beliefs. They are based on Koopman's initial justification of probability measure from qualitative requirements (Koopman, 1940, Fine, 1973). He introduces an order $\geq$ on the events where $A \geq B$ means that the probability of A is larger or equal to the probability of B. His major axiom is the Disjoint Union Axiom:

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

This axiom is central to the proof that the ordering relation is represented by a probability measure (Fine, 1973). Wong et al. (1990) propose to replace the Disjoint Union Axiom by a more general axiom.

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

In that case, they show that the ordering cannot always be represented by a probability measure, that it can always be represented by a belief function, a family of functions more general than the family of probability functions, but it can also be represented by other functions. Which of Koopman axiom or Wong et al. less demanding axiom must be satisfied is hardly obvious. They both seem reasonable but this cannot provide a definitive justification for their acceptation. Many other variants of these axioms have been suggested, but they always encountered the same problems. So whatever its beauty, the ordering justification is not compelling.

2.4. Cox algebraic justification.

Finally, Cox (1946) proposes an algebraic justification for the use of probability measure that has become quite popular in artificial intelligence domain (Cheeseman, 1985). He assumes that any measure of beliefs should essentially satisfy:

$$P(\bar{A}) = f(P(A)) \tag{2.1}$$

$$P(A \cap B) = F(P(A|B), P(B)) \tag{2.2}$$

for some f and F functions where f is decreasing in its argument. These requirements lead easily to the use of probability measure. But even the first requirement is criticized (Dubois et al., 1991). Suppose a medical context where a symptom is very frequent for a patient in disease class A and quite frequent for patient not in disease class A. Its observation should support the fact that the patient could belong to group A, this being translated by an increase in the probability that the patient belongs to group A. By (2.1) the fact that the patient belongs to group $\bar{A}$ becomes automatically less believable (in the sense of probable) even though the symptom is quite frequent in that group (but not as frequent as in group A). This unsatisfactory conclusion led some to reject (2.1) and (2.2). This was the origin of the certainty factor in artificial intelligence and developed in the 70's by Shortliffe and Buchanan (1975). This was also rejected in possibility theory (Zadeh, 1978, Dubois and Prade, 1985) and in Shafer's theory of evidence (Shafer, 1976a, Smets, 1978). Cox's axioms provide a nice characterization of probability measures. Nevertheless they are not really necessary.

2.5. Objective and subjective probabilities.

We deal with subjective probabilities but their link to objective probabilities must be considered. Hacking's Frequency Principle states that (Hacking, 1965):

> If the objective probability of A is p then the subjective probability of A is p.

It provides a very reasonable scale for Your subjective probability but it is based on the assumption that the objective probability exists. When You toss a coin and say that the probability of heads is 0.5, what is the 0.5 about? Hacking suggests it is a property of the chance set-up, i.e. the coin and the procedure that tosses the coin.... In that case, the .5 would preexist to the belief holder and a rational belief holder should use the .5 to quantify his belief about observing heads on the next toss. So in Hacking's perspective, there exists something called the objective probability and the subjective belief induced by a chance set-up and its related objective probability should share the same scale. Again, this does not justify that beliefs in general are quantified by probability measure. It only justifies the representation of beliefs by probability functions when they are induced by a chance set-up. Life is not just a bunch of chance set-ups, and the domain of our beliefs is much larger then those relate to some chance set-ups.

De Finetti takes a much more extreme position by rejecting any form of probability that is not subjective. For De Finetti, the .5 is not even a property of the chance set-up but it is a property of the belief holder. De Finetti goes as far as claiming that You are free to assign whatever probability You would like to heads under the provision that You allocate the remaining probability to tails. A .7 to heads is perfectly valid (if tails receive .3).

We will not further discuss of the existence of objective probability. It seems that this is an acknowledged property in quantum mechanics (d'Espagnat, 1976). We feel that Hacking's ideas of objective probability, chance set-up and his Frequency Principle are ideas that could be accepted, what we will do in our development. Nevertheless, it does not lead to the use of probability function to quantify all forms of beliefs.

2.6. Conclusion about the probabilistic models.

None of the justifications proposed for the use of probability functions to represent quantified beliefs leads to the necessity of their use. We present in this paper alternative models that will represent belief at the credal level, i.e. at the mental level where beliefs are only "entertained" outside of any consideration for some underlying decision. I feel I can have some belief about the weather in Hawaii...... even though I do not take any decision based on that weather. Degrees of belief can be assimilated to degree of conviction, degree of support, degree of justified belief.... (Voorbraak, 1993).

3. Non-Standard Probability Models.

We successively examined models based on upper and lower probabilities, sets of probability functions, Dempster-Shafer theory, second order probabilities and probabilities of provability. All these models have in common the existence of some additive measure that underlies somehow the agent's beliefs.

3.1. Static components.

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, for all $A\subseteq\Omega$, that the probability P(A) belong to an 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, 1990, Jaffray, 1992).

One obtains a similar model 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.

A special case of upper and lower probabilities has been described by Dempster (1967, 1968) and underlies most interpretations of **Dempster-Shafer theory**. He assumes the existence of a probability measure on a space X and a one to many mapping M from X to another space 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.

Example 1: Dempster model.
Suppose a space $X = \{x_1, x_2\}$ and $P_X(\{x_1\}) = .3$, $P_X(\{x_2\}) = .7$. Suppose a mapping M form X to Y where $Y = \{y_1, y_2\}$ and $M(x_1) = \{y_1\}$, $M(x_2) = \{y_1, y_2\}$. What can we say about the probability P_Y over the space Y. We know $P_Y(\{y_1\})$ is at least .3, and might be 1.0 if x_2 was indeed mapped onto y_1. We know $P_Y(\{y_2\})$ is at most .7 as the probability .3 given to x_1 cannot be given to y_2. Finally $P_Y(\{y_1,y_2\}) = 1$. Our knowledge about P_Y can be represented by the lower probability function P_{Y*}, with $P_{Y*}(\{y_1\}) = .3$, $P_{Y*}(\{y_1\}) = .7$, $P_{Y*}(\{y_1,y_2\}) = 1$. It happens that the lower probability function obtained through the M mapping is a belief function, what does not necessary mean that the lower probability function quantifies the agent's beliefs.

But what is a belief function? It is a Choquet capacity monotone of order infinite (Choquet, 1953) that Shafer (1976a) introduces in order to represent beliefs. Let Ω be a set and let 2^Ω be the Boolean algebra of subsets of Ω. A belief function is a function bel from 2^Ω to [0, 1] such that :

1) $bel(\emptyset) = 0$
2) $\forall A_1,A_2,...A_n \in \Re$,

$$\mathrm{bel}(A_1 \cup A_2 \cup \ldots A_n) \geq \sum_i \mathrm{bel}(A_i) - \sum_{i>j} \mathrm{bel}(A_i \cap A_j) \ldots\text{.-}(-1)^n \mathrm{bel}(A_1 \cap A_2 \cap \ldots A_n) \qquad (3.1)$$

Notice that we do not require that bel(Ω) = 1 as is usually accepted. We can have bel(Ω)<1, the difference 1-bel(Ω) quantifies the amount of internal conflict in the information that leads to the construction of bel (Smets, 1992a). This unnormalization of bel is not very important at this point and will not be further discussed.

Associated to every belief function bel, the is another function m from 2^Ω to [0, 1], called the basic belief assignment, which values m(A) for $A \subseteq \Omega$ are called the basic belief masses.

$$m(A) = \sum_{B:B\subseteq\Omega, \emptyset\neq B\subseteq A} (-1)^{|A|-|B|}.\ \mathrm{bel}(B) \qquad \forall A\subseteq\Omega,\ A\neq\emptyset$$

$$m(\emptyset) = 1 - \mathrm{bel}(\Omega)$$

and,
$$\mathrm{bel}(A) = \sum_{B:B\subseteq\Omega, \emptyset\neq B\subseteq A} m(B) \qquad \forall A\in\mathfrak{R},\ A\neq\emptyset$$

Coming back to the upper and lower probability models, their generalization into **second-order probability** models 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.

Ruspini (1986) and Pearl (1988) have suggested that the degree of belief given to a proposition that is quantified by a belief function could be understood as the probability that the proposition in know, is provable. Indeed, if one applies probability theory to modal propositions, one discovers easily that the ‘probability of proving’ is a belief function. This idea is also what Kohlas models in his hints models (Kohlas and Monney, 1994a). He assumes assumptions and hypothesis, knowing under which assumptions each hypothesis can be proved. These models are in fact part of the Dempster models, as each assumes a space endowed with a probability function and the provability relation plays the role of the one-to-many mapping M. The problems with these problems appear once conditioning is introduced. This will be further discussed in section 3.2, where the dynamic of beliefs is analyzed.

3.2. Dynamic components.

Suppose the agent had a given beliefs at time t_0 belief. At time $t_1 > t_0$, the agent learns for sure that ‘the proposition A is true’, or equivalently that ‘the actual world is in A’ for some A. We denote this information by Ev_A. We also suppose that, between t_0 and t_1, the agent has not learned anything relevant to his beliefs under consideration. So Ev_A is the first information relevant to the agent’ beliefs since t_0. This situation corresponds to the one described by the conditioning process in probability theory.

How does the agent change his beliefs once he learns Ev_A? This is the first question to be answered when the dynamic of beliefs is studied.

In probability theory, conditioning is achieved by Bayes rule of conditioning. In the upper and lower probability models, the second-order probability models and the family of probability functions models, the conditioning is performed for each probability function compatible with the constraints imposed by the static representation. A new set of probability functions is computed from these conditional probability functions. This method is quite immediate. In strict upper and lower probability models, one must be cautious as conditioning cannot be simply iterated. In fact, the knowledge of the upper and lower probability functions is not sufficient to represent a state of belief, one must replace them by upper and lower expectations in which case conditioning can be iterated (Walley, 1991).

With Dempster models as well as with the probability of provability models, conditioning becomes more delicate. The so-called Dempster's rule of conditioning proposed by Shafer is not blindly applicable. Suppose the agent learns that the real value of Y belongs to some $B \subseteq Y$. Does it means that the real value of X belongs to the set of $x \in X$ such $M(x) \subseteq B$? If it is the case, then the appropriate rule of conditioning is the so-called geometric rule (Shafer, 1976b). If the information means that the agent has learned a new information that states that the true value of Y belongs to B, and if this information means that the mapping M between X and Y must be revised into M_B such that for every $x \in X$, $M_B(x) = M(X) \cap B$, then Dempster's rule of conditioning is the appropriate rule (except for the problem of normalization).

In the probability of provability model, these two forms of conditioning correspond to:

case 1: the probability that the assumption selected randomly according to P_X will prove the hypothesis C given it proves the hypothesis B.
case 2: the probability that the assumption selected randomly according to P_X will prove the hypothesis C given You know that the hypothesis B has been proved by another pieces of evidence.

Once probabilities are generalized into non-standard probability models, conditioning can takes many forms, and the selection of the appropriate rules is delicate (Smets, 1991b). Most criticisms against Dempster-Shafer theories are essentially due to an inappropriate use of Dempster's rule of conditioning.

3.3. Decisions Making.

The most important quality of the Bayesian model is that is provides an excellent, if not the best tool, for optimal decision making under risk. But the decisions considered by Bayesians, an in particular in the Dutch Books arguments they favor so often, are all based on forced bets. The players must bet, and must be ready to be in either of the two positions, player or banker. In that case, betting without respecting the probability axioms would lead the agent to a sure loss, turning him into a money pump, a situation nobody could accept (see section 2.2). But the argument collapses once bets are neither forced nor symmetrical. Models based on upper and lower probabilities claim that the lower probability is the maximal prize the agent is ready to pay if he is the player, and the upper probability is the minimum prize the agent would ask from the player if the agent happens to be the banker. Such a model presents the weakness that alternatives are no more ordered as in the probability approach. Some pairs of decisions cannot be compared any more in order to decide which is best. Is it a good or a bad property is a matter of personal opinions.

3. 4. Conclusions about the non standard probability models.

The models answer to the questions that must be answered by any model for the representation of quantified beliefs (see section 1). They can even solve some of the weaknesses of the probability model. Total ignorance can be represented by just accepting that for any $A \subseteq \Omega$, $A \neq \Omega$, its lower probability is null. Such a state cannot be represented in pure probability model and the use of the Principle of Insufficient Reason is the best way to get trapped into paradoxes, inconsistencies, if not plain non-sense. As a matter of fact, Bayesians usually claim that a state of total ignorance does not exist, a preposterous attitude. If I ask You what are Your beliefs about the location of Pa’acal burial on Earth, I hardly think You could give any reasonable and justified answer, as far as You don’t even know who Pa'acal is[3]. Claiming total ignorance does not exist is nothing more than sweeping it under the carpet.

4. Non probability Models.

We analyze the two major non probability models based respectively on possibility functions and on belief functions.

4.1. Possibility Theory.

4.1.1. Possibility and Necessity Measures.

Incomplete 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. This can be represented by a 'possibility' measure defined on the height domain whose value is 0 if h < 170 and 1 if h is ≥ 170 (with 0 = impossible and 1 = possible). Ignorance results from the lack of precision, of specificity of the information "above 170".

When the predicate is vague like in 'John is tall', possibility can admit degrees, the largest the degree, the largest the possibility. But even though possibility is often associated with fuzziness, the fact that non fuzzy (crisp) events can admit different degrees of possibility is shown in the following example. Suppose there is a box in which you try to squeeze soft balls. 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 and they are totally unrelated to any supposedly underlying random process.

Identically ask a salesman about his forecast about next year sales. He could answer: it is possible to sell about 50K, impossible to sell more than 100K, quite possible to sell 70K, hardly possible to sell more than 90K... His statements express what are the possible values for next year sales. What the values express are essentially the sale capacity. Beside, he could also express his belief 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.

Let $\Pi:2^{\Omega}\rightarrow[0, 1]$ 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

[3]Pa'acal is the Mayan king burried at Palenque.

that the possibility $\Pi(A \vee B)$ of the disjunction of two propositions A and B is the maximum of the possibility of the individual propositions $\Pi(A)$ and $\Pi(B)$. (Zadeh, 1978, Dubois and Prade, 1985):

$$\Pi(A \vee B) = \max (\Pi(A) , \Pi(B)). \quad (4.1)$$

Usually one also requires $\Pi(\Omega) = 1$.

As in modal logic, where the **necessity** of a proposition is the negation of the possibility of its negation, one defines the necessity measure N(A) given to a proposition A by:

$$N(A) = 1 - \Pi(\neg A)$$

In that case, one has the following:

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

Beware that one has only:

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

$$N(A \vee B) \geq \max (N(A) , N(B)).$$

Let Ω be the universe of discourse on which a possibility measure Π is defined. Related to the possibility measure $\Pi : 2^{\Omega} \rightarrow [0, 1]$, one can define a **possibility distribution** $\pi : \Omega \rightarrow [0, 1]$,

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

Thanks to (4.1), one has

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

A very important point in possibility theory (and in fuzzy set theory) when only the max and min operators are used is the fact that the values given to the possibility measure (or to the grade of membership) are not intrinsically essential. The only important element of the measure is the order they create among the elements of the domain. Indeed the orders are invariant under any strictly monotonous transformation. Therefore a change of scale will not affect conclusions. This property explains why authors insist on the fact that possibility theory is essentially an ordinal theory, a nice property in general. This robustness property does not apply once addition and multiplication are introduced as is the case with probability and belief functions, or when operators different from the min-max operators are used.

Example 2: Probability versus possibility. Hans eggs.

As an example of the use of possibility measure versus probability measure, consider the number of eggs X 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. 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 1.

Table 4: The possibility and probability distributions associated with X.

u	1	2	3	4	5	6	7	8
π(u)	1	1	1	1	.8	.6	.4	.2
p(u)	.1	.8	.1	0	0	0	0	0

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).

4.1.2. Physical and Epistemic Possibility.

Two forms of (continuous valued) possibility have been described: the physical and the epistemic. These 2 forms of possibility can be recognized by their different linguistic uses: it is **possible that** and it is **possible for** (Hacking, 1975). When I say it is possible that Paul's height is 170, it means that for all I know, Paul's height may be 170. When I say it is possible for Paul's height to be 170, it means that physically, Paul's height may be 170. The first form, 'possible that', is related to our state of knowledge and is called epistemic. The second form, 'possible for', deals with actual abilities independently of our knowledge about them. It is a degree of realizability. The distinction is not unrelated to the one between the epistemic concept of probability (called here the credibility) and the aleatory one (called here chance). These forms of possibilities are evidently not independent concepts, but the exact structure of their interrelations is not yet clearly established.

4.1.3. Relation between fuzziness and possibility.

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

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

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

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

This writing of Zadeh's possibilistic principle is the one most usually encountered but its meaning should be interpreted with care. 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. The writing is often confusing and would have been better written as

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

or still better

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

The last expression avoids the confusion between the two concepts. Just as with Hacking Frequency Principle, 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 between μ and π 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) in which case Zadeh's possibilistic principle becomes the likelihood principle:

$$l(\theta|x) = p(x|\theta)$$

The likelihood of an hypothesis θ given an observation x is numerically equal to the probability of the observation x given the hypothesis θ.

4.2. The Transferable Belief Model.

4.2.1. Static and Dynamic Representations.

We are now exploring a model for representing quantified beliefs based on belief functions, the transferable belief model. No concept of randomness, or probability, is involved. We want to study the appropriate model that should be used to represent beliefs at the credal level. When randomness is not involved, there is no necessity for beliefs at the credal states (the psychological level where beliefs are entertained) to be quantified by probability measures (Levi, 1984). The coherence principle advanced by the Bayesians to justify probability measures is adequate in a context of decision (Degroot, 1970), but it cannot be used when all we want to describe is a cognitive process. Beliefs can be entertained outside any decision context. In the transferable belief model (Smets, 1988) we assume that beliefs at the credal level are quantified by belief functions (Shafer, 1976a). When decisions must be made, our belief held at the credal level induces a probability measure held at the so-called 'pignistic' level (the level at which decisions are made). This probability measure will be used in order to make decisions using expected utilities theory. But it is important to stress that this probability measure is not a representation of our belief, but is only induced from it when decision is involved.

The next example illustrate the concept we want to cover with the transferable belief model.

Example 3. Let us consider a somehow reliable witness in a murder case who testifies to You that the killer is a male. Let $\alpha = .7$ be the reliability You give to the testimony. Suppose that *a priori* You have an equal belief that the killer is a male or a female. A classical probability analysis would compute the probability P(M) of M = 'the killer is a male'. P(M) = .85 = .7 + .5 x.3 (the probability that the witness is reliable (.7) plus the probability of M given the witness is not reliable (.5) weighted by the probability that the witness is not reliable (.3)). The transferable belief model analysis will give a belief .7 to M: bel(M) = .7. In P(M) = .7 + .15, the .7 value can be viewed as the *justified* component of the probability given to M (called the belief or the support) whereas the .15 value can be viewed as the *aleatory* component of that probability. The transferable belief model deals only with the justified components. (Note: the Evidentiary Value Model (Ekelof, 1982, Gärdenfors et al., 1983) describes the same belief component, but within a strict probability framework, and differs thus from the transferable belief model once conditioning is introduced.)

Further suppose there are only two potential male suspects: Phil and Tom. Then You learn that Phil is not the killer. The testimony now supports that the killer is Tom. The reliability .7 You gave to the testimony initially supported 'the killer is Phil or Tom'. The new information about Phil implies that .7 now supports 'the killer is Tom'. □

Hence the major component of the transferable belief model are the parts of beliefs that supports some proposition without supporting any strictly more specific propositions. These parts of beliefs are called the basic belief masses. A basic belief mass given to a set A supports also that the actual world is in every subsets that contains A. The degree of belief bel(A) for $A \in \Re$ quantifies the total amount of *justified specific support* given to A. It is obtained by summing all basic belief masses given to subsets $X \in \Re$ with $X \subseteq A$ (and $X \neq \emptyset$)

$$\mathrm{bel}(A) = \sum_{\emptyset \neq X \subseteq A, X \in \Re} m(X)$$

We say *justified* because we include in bel(A) *only* the basic belief masses given to subsets of A. For instance, consider two distinct atoms x and y of $\Re$. The basic belief mass m({x,y}) given to {x,y} could support x if further information indicates this. However given the available information the basic belief mass can only be given to {x,y}. In example 3, the .7 was given to the set {Phil, Tom} and not split among its elements. We say *specific* because the basic belief mass m(Ø) is not included in bel(A) as it is given to the subset Ø that supports not only A but also $\bar{A}$.

Given a belief function bel, we can define a dual function that formalizes the concept of plausibility. The degree of plausibility pl(A) for $A \in \Re$ quantifies the maximum amount of *potential specific support* that could be given to A. It is obtained by adding all those basic belief masses given to subsets X compatible with A, i.e., such that $X \cap A \neq \emptyset$:

$$pl(A) = \sum_{X \cap A \neq \emptyset, X \in \Re} m(X) = bel(\Omega) - bel(\bar{A})$$

We say *potential* because the basic belief masses included in pl(A) could be transferred to non-empty subsets of A if new information could justify such a transfer. It would be the case if we learn that $\bar{A}$ is impossible.

The function pl is called a plausibility function. It is in one-to-one correspondence with belief functions. It is just another way of presenting the same information and could be forgotten, except inasmuch as it often provides a mathematically convenient alternate representation of our beliefs.

If some further evidence becomes available to You and implies that B is true, then the mass m(A) initially allocated to A is transferred to $A \cap B$. This transfer of the basic belief masses characterizes the conditioning process described in the transferable belief model. It is called the Dempster's rule of conditioning, and provides the major element to describe the dynamic of beliefs.

Total ignorance is represented by a vacuous belief function, i.e. a belief function such that $m(\Omega) = 1$, hence $bel(A) = 0 \ \forall A \in \Re$, $A \neq \Omega$, and $bel(\Omega) = 1$. The origin of this particular quantification for representing a state of total ignorance can be justified. Suppose that there are three propositions labeled A, B and C, and You are in a state of total ignorance about which is true. You only know that one and only one of them is true but even their content is unknown to You. You only know their number and their label. Then You have no reason to believe any one more than any other, hence, Your beliefs about their truth are equal: $bel(A) = bel(B) = bel(C) = \alpha$ for some $\alpha \in [0,1]$. Furthermore, You have no reason to put more or less belief in $A \cup B$ than in C: $bel(A \cup B) = bel(C) = \alpha$

(and similarly bel($A \cup C$) = bel($B \cup C$) = α). The vacuous belief function is the only belief function that satisfies equalities like: bel($A \cup B$) = bel(A) = bel(B) = α. Indeed the inequalities (3.1) are such that bel($A \cup B$) $\geq$ bel(A) + bel(B) - bel($A \cap B$). As $A \cap B = \emptyset$, bel($A \cap B$) = 0. The inequality becomes $\alpha \geq 2\alpha$ where $\alpha \in [0,1]$, hence $\alpha = 0$. The basic belief assignment related by the vacuous belief function is called a vacuous basic belief assignment.

In general, the basic belief assignment looks similar to a probability distribution function defined on the power set 2^{Ω} of the frame of discernment Ω. This analogy led several authors to claim that the transferable belief model is nothing but a probabilistic model on 2^{Ω}. Such an interpretation does not resists once conditioning is introduced, as far as it does not lead to Dempster's rule of conditioning we derive in section 4.4 (Smets, 1992b).

Even though the conditioning process is by far the most important form of belief dynamic, other forms of belief dynamic have been developed within the transferable belief model. Most famous is the Dempster's rule of combination rules that allows us to combine conjunctively the belief functions induced by two distinct pieces of evidence. The disjunctive combination rule has been studied in Smets (1993b). Cautious rules applicable when the two pieces of evidence are not distinct are being developed. Generalization of the Bayesian theorem, so important in any inferential procedure within the Bayesian approach, has also been developed and justified in Smets (1993b).

4.2.2. Decision Making.

When a decision must be made that depends on the actual world, the agent constructs a probability function in order to make the optimal decision, i.e., the one that maximizes the expected utility (Savage, 1954, DeGroot, 1970). As far as beliefs guide our actions, the probability function is a function of the belief function bel that describes the agent's belief at the credal level. Hence one must transform bel into a probability function, denoted BetP. This transformation is called the pignistic transformation. We call BetP a pignistic probability to insist on the fact that it is a probability measure used to make decisions (Bet is for betting). Of course BetP is a classical probability measure.

The structure of the pignistic transformation is derived from and justified by the following scenario.

Example 4: Buying Your friend's drink. Suppose You have two friends, G and J. You know they will toss a fair coin and the winner will visit You tonight. You want to buy the drink Your friend would like to have tonight: coke, wine or beer. You can only buy one drink. Let D = {coke, wine, beer}.

Let $bel_G(d)$, for all $d \subseteq D$, quantifies Your belief about the drink G is liable to ask for. Given bel_G, You build the pignistic probability $BetP_G$ about the drink G will ask by applying the (still to be defined) pignistic transformation. You build in identically the same way the pignistic probability $BetP_J$ based on bel_J, Your belief about the drink J is liable to ask for. The two pignistic probability distributions $BetP_G$ and $BetP_J$ are the conditional probability distributions about the drink that will be asked for given G or J comes. The pignistic probability distributions $BetP_{GJ}$ about the drink that Your visitor will ask for is then:

$$BetP_{GJ}(d) = .5\ BetP_G(d) + .5\ BetP_J(d) \qquad \text{for all } d \in D$$

You will use these pignistic probabilities $BetP_{GJ}(d)$ to decide which drink to buy.

But You might as well reconsider the whole problem and first compute Your belief about the drink Your visitor (V) would like to have. It can be proved that:

$$bel_V(d) = .5\ bel_G(d) + .5\ bel_J(d) \qquad \text{for all } d \subseteq D$$

Given bel_V, You could then build the pignistic probability $BetP_V$ You should use to decide which drink to buy. It seems reasonable to assume that $BetP_V$ and $BetP_{GJ}$ must be equal. In such a case, the pignistic transformation is uniquely defined. □

Given a belief function defined on Ω, its pignistic transformation BetP is:

$$BetP(\omega) = \sum_{A:\omega \in A \subseteq \Omega} \frac{m(A)}{|A|\ (1-m(\emptyset))} \qquad \text{for } \omega \in \Omega. \tag{4.2}$$

and $$BetP(A) = \sum_{\omega \in A} BetP(\omega)$$

where $|A|$ is the number of atoms of $\Re$ in A (Smets, 1990b, Smets and Kennes, 1994).

It is easy to show that the function BetP is a probability function and the pignistic transformation of a probability function is the probability function itself.

Historical note. In a context close to ours, Shapley (1953) derived relation (4.2). Amazingly, the model he derived was called the 'transferable utility model' whereas, independently, we called our model the 'transferable belief model'.

4.2.3. The impact of the two-level model.

In order to show that the introduction of the two-level mental model based on the credal and the pignistic levels, is not innocuous, we present an example where the results will be different if one takes the two-level approach as advocated in the transferable belief model or a one-level model like in probability theory.

Example 5: The Peter, Paul and Mary Saga.

Big Boss has decided that Mr. Jones must be murdered by one of the three people present in his waiting room and whose names are Peter, Paul and Mary. Big Boss has decided that the killer on duty will be selected by a throw of a dice: if it is an even number, the killer will be female, if it is an odd number, the killer will be male. You, the judge, know that Mr. Jones has been murdered and who was in the waiting room. You know about the dice throwing, but You do not know what the outcome was and who was actually selected. You are also ignorant as to how Big Boss would have decided between Peter and Paul in the case of an odd number being observed. Given the available information at time t_0, Your odds for betting on the sex of the killer would be 1 to 1 for male versus female.

At time $t_1 > t_0$, You learn that if Big Boss had not selected Peter, then Peter would necessarily have gone to the police station at the time of the killing in order to have a perfect alibi. Peter indeed went to the police station, so he is not the killer. The question is how You would bet now on male versus female: should Your odds be 1 to 1 (as in the transferable belief model) or 1 to 2 (as in the most natural Bayesian model).

Note that the alibi evidence makes 'Peter is not the killer' and 'Peter has a perfect alibi' equivalent. The more classical evidence 'Peter has a perfect alibi' would only imply P('Peter is not the killer' | 'Peter has a perfect alibi') = 1. But P('Peter has a perfect alibi' | 'Peter is not the killer') would stay undefined and would then give rise to further discussion, which would be useless for our purpose. In this presentation, the latter probability is also 1.

The transferable belief model solution.

Let k be the killer. The information about the waiting room and the dice throwing pattern induces the following basic belief assignment m_0:

$$k \in \Omega = \{\text{Peter, Paul, Mary}\}$$

$$m_0(\ \{\text{Mary}\}\ ) = .5 \qquad\qquad m_0(\ \{\text{Peter, Paul}\}\ ) = .5$$

The .5 belief mass given to {Peter, Paul} corresponds to that part of belief that supports "Peter or Paul", could possibly support each of them, but given the lack of further information, cannot be divided more specifically between Peter and Paul.

Let BetP_0 be the pignistic probability obtained by applying the pignistic transformation to m_0 on the betting frame which set of atoms is {{Peter], {Paul}, {Mary}}. By relation (4.2), we get:

$$\text{BetP}_0(\{\text{Peter}\}) = .25 \qquad \text{BetP}_0(\{\text{Paul}\}) = .25 \qquad \text{BetP}_0(\{\text{Mary}\}) = .50$$

Given the information available at time t_0, the bet on the killer's sex (male versus female) is held at odds 1 to 1.

Peter's alibi induces an updating of m_0 into m_2 be Dempster's rule of conditioning:

$$m_2(\{Mary\}) = m_2(\{Paul\}) = .5$$

The basic belief mass that was given to "Peter or Paul" is transferred to Paul.

Let $BetP_2$ be the pignistic probability obtained by applying the pignistic transformation to m_2 on the betting frame which set of atoms is {{Paul}, {Mary}}.

$$BetP_2(\{Paul\}) = .50 \qquad BetP_2(\{Mary\}) = .50$$

Your odds for betting on Male versus Female would still be 1 to 1.

The probabilistic solution:

The probabilistic solution is not obvious as one data is missing: the value α for the probability that Big Boss selects Peter if he must select a male killer. Any value could be accepted for α, but given the total ignorance in which we are about this value, let us assume that α =.5, the most natural solution. Then the odds on male versus female before learning about Peter's alibi is 1 to 1, and after learning about Peter's alibi, it becomes 2 to 1. The probabilities are then:

$$P_2(\{Paul\}) = .33 \qquad P_2(\{Mary\}) = .66$$

The 1 to 1 odds of the transferable belief model solution can only be obtained in a probabilistic approach if $\alpha = 0$. Some critics would claim that the transferable belief model solution is valid as it fits with $\alpha = 0$. The only trouble with this answer is that if the alibi story had applied to Paul, than we would still bet at 1 to 1 odds. Instead the probabilistic solution with $\alpha = 0$ would lead to a 0 to 1 bet, as the probabilities are:

$$P_2(\{Peter\}) = .0 \quad P_2(\{Mary\}) = 1.$$

So the classical probabilistic analysis does not lead to the transferable belief model solution.

We are facing two solutions for the bet on male versus female after learning about Peter's alibi: the 1 to 1 or at 1 to 2 odds? Which solution is 'good' is not decidable, as it would require the definition of 'good'. Computer simulations have been suggested for solving the dilemma, but they are impossible. Indeed for every simulation where the killer must be a male, one must select Peter or Paul, and as far as simulation are always finite, the proportion of cases when Peter was selected in the simulations when a male has to be selected will be well defined. The value of the two solutions under comparison will only reflect the difference between that proportion and the missing probability α. Such a comparison is irrelevant for what we look for. We are only left over with a subjective comparison of the two solution... or an in depth comparison of the theoretical foundations that led to these solutions, an alternative that explains why we try to develop a full axiomatization of the transferable belief model (Smets, 1993c)

Example 6: The Five Breakable Sensors.
To show that the choice between a Bayesian and a transferable belief model approach is important in practice, we analyze the story of the five breakable sensors where the two approaches completely disagree in their conclusions, and would led to acting differently. These examples might help the reader in deciding which approach is better.

Suppose I must check the temperature of a process. To do it I have five sensors, and I have the same information for each of them.

Each sensor can check the temperature of the process. The temperature can only be hot or cold. If the temperature is hot (TH), the sensor light is red (R) and if the temperature is cold (TC), the sensor light is blue (B). Each sensor is made of a thermometer and a device that turns the blue or the red light on according to the temperature reading. Unfortunately the thermometer can be broken.

The only known information is what is written on the box containing each sensor. "Warning: the thermometer included in this sensor can be broken. The probability that it is broken is ...%. (a number is written there that depends on the box). When the thermometer is not broken, the sensor is a perfectly reliable detector of the temperature. When the thermometer is not broken, red light means the temperature is hot, blue light means that the temperature is cold. When the thermometer is broken, the sensor answer is unrelated to the temperature".

I am a new technician and I never saw the five sensors before. On box 1, the probability is 1% that the thermometer is broken, on the boxes 2 to 5, the probability is 35%. I know nothing about these sensors except the warnings written on the boxes. I use them and the red light gets on for the box 1 sensor, and the blue light gets on for boxes 2 to 5 sensors. How do I make up my mind about the temperature status? What is my opinion (belief) that the temperature status is hot or cold? Let us assume that the consequences (utilities) of the good and bad decisions are symmetrical: I am either right or wrong. We can thus avoid discussions about possible consequences. The problem is: does the agent belief that the temperature is hot increases or decreases given the five observations. With the transferable belief model, the probability increases, with the Bayesian analysis it decreases. Bayesians could conclude that the temperature is cold. The transferable belief model could concludes that the temperature is hot. To decide which is more natural is left to the reader!

The lesson of the examples 5 and 6 is that the choice of the model is important and interferes with real decisions.

5. Conclusions.

No definitive conclusions can be taken from our presentation, except that uncertainty can take several forms, and that the choice of the appropriate model is a necessity. Any claim like ‘my model is the best and the only valid model for representing quantified belief’ is just nonsensical. It characterized a dogmatic approach without any relation with a scientific attitude.

We did not tackle the problem of assessing the values of the belief. This problem is well solved in probability theory, thanks to the exchangeable bets schema. The same solution is used within the transferable belief model. For upper and lower probability models, the assessment is not so well defined as forced symmetrical bets are not available. The assessment problem is not acute in ordinal possibility theory, i.e., the one where only the ordering is relevant, not the value themselves. In cardinal possibility theory, the problem is not yet resolved.

The problem of finding which is the appropriate model for which model is hardly solved, and justifies further researches. We hope we have provided some tools in that direction by presenting the various models that have been so far proposed for the quantified representation of beliefs.

Bibliography.

CHEESEMAN P. (1985) In defense of probability. IJCAI-85, Morgan Kaufman, San Mateo, Ca, 1002-1009.

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.

COHEN L. J., (1993) What has probability to do with strengh of belief. in DUBUCS J.P. (ed.) Philosophy of Probability, Kluwer, Dordrecht, pg. 129-143.

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

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. A generalization of Bayesian inference. J. Roy. Statist. Soc. B.30 (1968) 205-247.

D’ESPAGNAT B. (1976) Conceptual foundations of quantum mechanics. W.A. Benjamin, Reading, Mass.

DUBOIS D., GARBOLINO P., KYBURG H.E., PRADE H. and SMETS P. (1991) Quantified Uncertainty. J. Applied Non-Classical Logics 1:105-197.

DUBOIS D. and PRADE H. (1985) *Theorie des possibilités.* Masson, Paris.

EKELOF P.O. (1982) Rättegång IV. Fifth edition, Stockholm.

FAGIN R. and HALPERN J. (1990) A new approach to updating beliefs. 6th Conf. on Uncertainty in AI.

FINE T. (1973) Theories of probability. Academic Press, New York.

GÄRDENFORS P., HANSSON B. and SAHLIN N.E. (1983) Evidentiary value: philosophical, judicial and psychological aspects of a theory. C.W.K. Gleerups, Lund.

GÄRDENFORS P. (1988) Knowledge in flux. Modelling the dynamics of epistemic states. MIT Press, Cambridge, Mass.

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.

HACKING I. (1965) Logic of statistical inference. Cambridge University Press, Cambridge, U.K.

HACKING I.(1975) The emergence of probability. Cambridge University Press, Cambridge.

HINTIKKA J. (1962) Knowledge and belief. Cornell Univ. Press, Ithaca, NY.

JAFFRAY J.Y. (1989) Coherent bets under partially resolving uncertainty and belief functions. Theory and Decision 26:99-105.

JAFFRAY J.Y. (1992) Bayesian updating and belief functions. IEEE Trans. SMC, 22: 1144-1152.

KOHLAS J. and MONNEY P.A (1994a) Representation of Evidence by Hints. in Advances in the Dempster-Shafer Theory of Evidence. Yager R.R., Kacprzyk J. and Fedrizzi M., eds, Wiley, New York, pg. 473-492.

KOHLAS J. and MONNEY P.A. (1994b) Theory of Evidence. A Survey of its Mathematical Foundations, Applications and Computations. ZOR-Mathematical Methods of Operatioanl Research 39:35-68.

KOOPMAN B.O. (1940) The bases of probability. Bull. Amer. Math. Soc. 46:763-774.

KYBURG H. (1987) Objective probabilities. IJCAI-87, 902-904.

LEVI I. (1984) Decisions and revisions: philosophical essays on knowledge and value. Cambridge University Press, Cambridge.

PEARL J. (1988) Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann Pub. San Mateo, Ca, USA.

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.

RUSPINI E.H. (1986) The logical foundations of evidential reasoning. Technical note 408, SRI International, Menlo Park, Ca.

SAHLIN N. E. (1993) On higher order beliefs. in DUBUCS J.P. (ed.) Philosophy of Probability, Kluwer, Dordrecht, pg. 13-34.

SAVAGE L.J. (1954) Foundations of Statistics. Wiley, New York.

SHAFER G. (1976a) 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. (1978) Nonadditive probabilities in the work of Bernoulli and Lambert. Arch. History Exact Sci. 19:309-370.

SHAPLEY L.S. (1953) A value for n-person games. In Contributions to the Theory of Games, vol. 2, eds. H. Kuhn and A.W. Tucker. Princeton University Press, pp. 307-317.

SHORTLIFFE E.H. and BUCHANAN B.G. (1975) A model of inexact reasoning in medicine. Math. Biosci. 23:351-379.

SMETS P. (1978) Un modèle mathématico-statistique simulant le processus du diagnostic médical. Doctoral dissertation, Université Libre de Bruxelles, Bruxelles, (Available through University Microfilm International, 30-32 Mortimer Street, London W1N 7RA, thesis 80-70,003)

SMETS P. (1987) Upper and lower probability functions versus belief functions. Proc. International Symposium on Fuzzy Systems and Knowledge Engineering, Guangzhou, China, July 10-16, pg 17-21.

SMETS P. (1988) Belief functions. in SMETS Ph, MAMDANI A., DUBOIS D. and PRADE H. ed. Non standard logics for automated reasoning. Academic Press, London pg 253-286.

SMETS P. (1990a) The combination of evidence in the transferable belief model. IEEE-Pattern analysis and Machine Intelligence, 12:447-458.

SMETS P. (1990b) 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 P. (1991a) Varieties of ignorance. Information Sciences. 57-58:135-144.

SMETS P. (1991b) About updating. in D'Ambrosio B., Smets P., and Bonissone P.P. eds, Uncertainty in AI 91, Morgan Kaufmann, San Mateo, Ca, USA, 1991, 378-385.

SMETS P. (1991c) Probability of provability and belief functions. Logique et Analyse, 133-134:177-195.

SMETS P. (1992a) The nature of the unnormalized beliefs encountered in the transferable belief model. in Dubois D., Wellman M.P., d’Ambrosio B. and Smets P. Uncertainty in AI 92. Morgan Kaufmann, San Mateo, Ca, USA, 1992, pg.292-297.

SMETS P. (1992b) The transferable belief model and random sets. Int. J. Intell. Systems 7:37-46.

SMETS P. (1993a) Imperfect information: imprecision - uncertainty. in Smets P. and Motro A. eds. Uncertainty Management in information systems: from needs to solutions. UMIS Workshop, Puerto Andraix.Volume 2. pg. 1-28.

SMETS P. (1993b) Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem. Int. J. Approximate Reasoning 9:1-35.

SMETS P. (1993c) An axiomatic justifiaction for the use of belief function to quantify beliefs. IJCAI'93 (Inter. Joint Conf. on AI), San Mateo, Ca, pg. 598-603.

SMETS P. (1994a) The representation of quantified belief by belief functions: an axiomatic justificiation. TR/IRIDIA/94.3. Submitted for publication.

SMETS P. (1994b) 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 P. and KENNES R. (1994) The transferable belief model. Artificial Intelligence 66:191-234.

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.

SMITH P. and JONES O.R. (1986) The philosophy of mind, an introduction. Cambridge University Press, Cambridge.

SMITHSON M. (1989) Ignorance and Uncertainty: Emerging Paradigms. Springer-Verlag, New York.

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.

WILLIAMS P.M. (1976) Indetermiinate probabilities. in Formal methods and the methodology of empirical sciences, Przlecki M., Szaniawski K. and Wojciki R. (eds.) Reidel, Dordrecht.

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.

ZADEH L.A. (1965) Fuzzy sets. Inform.Control. 8:338-353.

ZADEH L. (1978) Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems 1: 3-28.