PROBABILTY, POSSIBILITY, BELIEF: WHICH AND WHERE?

Philippe SMETS
IRIDIA, Université Libre de Bruxelles
50 av. Roosevelt, 1050 Brussels, Belgium
E-mail: psmets@ulb.ac.be

ABSTRACT

We analyze the difference between imprecision and uncertainty, trying to clarify their difference, in particular when possibility theory is introduced. We compare the applicablity of the various models for uncertainty.

1. Introduction.

The lack of a unique model to represent quantified uncertainty constitutes a problem for the user: which model should be applied in what situation ? We present some models for the quantified representation of uncertainty, focusing on their applicability more than on their mathematical structure. Imprecision often underlies uncertainty, hence we will also study imprecision.

We define imprecision and uncertainty as follows. There is imprecision whenever the exact value of the truth status of a proposition of interest is not established uniquely, i.e., whenever its truth status is equivocal. Uncertainty is an added information that expresses the idea that the truth of some propositions is better 'supported' than the truth of others. The paper will focus on trying to explain this distinction that is often confusing, and confused.

1.1. The actual world.

The models we are going to study quantify the uncertainty held by an agent about which possible world is the actual world. The general framework we use is based on the idea of propositions and possible worlds, even though the presentation could also be focusing on events, subsets, etc..... We consider that there is an actual world, denoted by ω_0, i.e., a world that corresponds to the actual state of affairs. The agent, denoted by You, does not know which world in a set of possible worlds is the actual world. This ignorance results from Your limited understanding (please, do not take it personally), from Your ignorance about the truth status of some propositions in the actual world. Should You know the truth status of every propositions of interest in ω_0, then you would know which world is ω_0.

Formally, we suppose a propositional language L, supplemented by the tautology and the contradiction, denoted $\top$ and $\bot$, respectively. The set of propositions will be finite even though the presentation could be adapted for countable or even uncountable sets of propositions. Let Ω_L denote the set of worlds that correspond to the interpretations of L and built so that no two worlds denote logically equivalent propositions, i.e., for every pair of worlds in Ω_L, there exists a proposition in the language L that is true in one world and false in the other (this avoids useless repetition of logically equivalent worlds). Among those worlds in Ω_L, a particular one, denoted ω_0, corresponds to the actual world. You ignore which world in Ω_L is ω_0. All You can express is Your 'opinion' about the fact that the world ω_0 belongs to the various subsets of Ω_L. This opinion can be representing some belief, probability, possibility, etc.... The list is hardly limited and we will discuss hereafter about various forms of 'opinions' that can be described. I will present personal opinions about these 'opinions' knowing that I might be biased, incomplete or, worse, wrong. But the nature of these 'opinions' is hardly well established and hoping for a perfect definition is just unrealistic.

1.2. The evidential corpus.

The set Ω_L is built independently of You, the agent. Your knowledge about which world in Ω_L is the actual world ω_0 is encoded in what we call the evidential corpus, denoted EC, where EC characterizes Your personal knowledge at a given time. The intended meaning of EC is 'all You know', i.e., the background knowledge on which You build Your opinions. It is not easy to formalize EC. Let 2^L denote the Boolean algebra of propositions built from L. Then EC contains a set K of propositions of 2^L known to be true to You. K is assumed to be consistent and deductively closed. Besides, EC also contains Your 'opinions' about the truth value in the actual world of those propositions of 2^L not in K.

That EC might and does contain more than what we have just put in it is not essential to our presentation. The only question is to determine how Your 'opinions' can be represented in order that EC would be somehow 'consistent'. Consistency is well defined in classical propositional logic, hence for those propositions in K. But when it comes to the 'opinions', what means consistency becomes unclear. For instance, Bayesians require that in order to be consistent, 'opinions' should be represented by additive weights, the probabilities, such that, among others, the probability $P(\phi)$ given to the fact that a proposition ϕ in 2^L is true in the actual world and the probability $P(\neg\phi)$ given to the fact that a proposition ϕ is false in the actual world add to one, etc... The origin of their requirements is based on the concepts of rational decision making and Dutch Books avoidance (see hereafter). Other proposals for consistency could lead to other models, as it is the case with the u-possibility theory [a],

[a]We introduce the notation u-possibility and i-possibility to distinguish between the possibility theory that deal with uncertainty (u-possibility) and with imprecision (i-possibility). Both are epistemic constructs, but concerns different, even though related, problems.

the transferable belief model, etc...

2. Imprecision.

Given the set K of propositions in EC, You can construct the set Ω_K of worlds where all the propositions in K are true:

$$\Omega_K = \{\omega : \omega \in \Omega_L, \omega \models \phi, \ \forall \phi \in K\}.$$

For what concerns You, Your universe is limited to Ω_K as You know that the worlds in Ω_L not in Ω_K do not satisfy Your knowledge encoded in K.

For ϕ in 2^L, we denote by $[\phi]$ the set of world in Ω_K where ϕ is true:

$$[\phi] = \{\omega : \omega \in \Omega_K, \omega \models \phi, \phi \in 2^L\}$$

Notice that $[\phi]$ is defined relatively to Ω_K and not Ω_L. This reflects that for what concerns You, only those worlds in Ω_K are relevant as You are 'sure' that ω_0 belongs to Ω_K (which might be wrong of course, but an agent may be wrong, which is not a good idea but reflects personal freedom).

So for all ϕ in K, You know '$\omega_0 \models \phi$'. For all ϕ such that $\neg\phi$ belongs to K, You know '$\omega_0 \models \neg\phi$'. If neither ϕ nor $\neg\phi$ belong to K, You know '$\omega_0 \models \phi$ or $\omega_0 \models \neg\phi$' (because of the excluded middle principle) but You do not know if '$\omega_0 \models \phi$' and You do not know if '$\omega_0 \models \neg\phi$'.

If all You know stops there, all You can state is that ω_0 belongs to Ω_K and if $|\Omega_K|$, the number of worlds of Ω_L in Ω_K, is larger than one (as it is usually the case) Your knowledge about ω_0 can be qualified as imprecise. At first sight, we could state that the largest $|\Omega_K|$ the largest the imprecision. It is wrong: the concept of imprecision is not related to $|\Omega_K|$ (once $|\Omega_K| > 1$). Imprecision pertain only to the question under consideration, where a question is a proposition for which You try to assess its truth status in ω_0. There are questions for which You know the answers (those where the answer is the same in every world in Ω_K, hence in ω_0, whatever ω_0 is actually) and others that can admit several answers (those where the answer varies among the worlds in Ω_K). The first category is hardly interesting as far as the answer to the question is obvious. The problem resides in the second case. In the first case we will say that 'ϕ is true in ω_0' is necessary (as well as it is possible), as it is true in every world in Ω_K and ω_0 belongs to Ω_K. In the second case we say that 'ϕ is true in ω_0' is possible (but not necessary), as there are worlds in Ω_K where ϕ is true and other where ϕ is false. In the first case we say that ϕ is necessary and possible, in which case $\neg\phi$ is impossible. In the second we say ϕ is possible but not necessary, in which case $\neg\phi$ is also possible but not necessary. As a shortcut, we will say that (given K and for You), ϕ is necessary in the first case and contingent in the second case. Table 1 presents the four classical modalities, with their classical notation where $\Box$ denotes

Necessary and $\Diamond$ denotes Possible. Their link with the universal $\forall$ and existensial $\exists$ quantifiers are also indicated.

Necessary	$\Box$	$\neg\Diamond\neg$	$\forall x$
Contingent	$\neg\Box$	$\Diamond\neg$	$\neg\forall x$
Possible	$\neg\Box\neg$	$\Diamond$	$\exists x$
Impossible	$\Box\neg$	$\neg\Diamond$	$\neg\exists x$

Table 1: The four modalities.

Suppose a contingent proposition ϕ (a question) for which You want to assess the truth status in ω_0. What can You say about the truth status of ϕ in ω_0? Up to here, nothing more than it might be true or it might be false. You could start counting the numbers of worlds where ϕ is true and those where ϕ is false, from which You can compute the proportion p of worlds where ϕ is true (with $p = |[\phi]|/|\Omega_K|$). Except in the limiting cases where $p = 1$ or $p = 0$, in which case ϕ is trivially known to be true or known to be false, respectively, You can say nothing except that ϕ is true in a proportion p of worlds in Ω_K, which should not induce You in believing that ϕ is true with a degree p. Proportion in itself has nothing to do with beliefs, except in the extreme cases. Proportions might become relevant to beliefs once 'equiprobability' or similar extra properties are introduced in EC.

Besides, the meaning of that proportion p is defined relatively to L. If we change the propositional space L into another propositional space L^*, usually the proportion will also change. [b]We had assumed some initial language L, but we could reconsider the whole analysis by using another language L^*, and we feel that the opinion should not be influenced by the nature of L as far as L is 'rich' enough to express Your knowledge (whatever 'rich' means exactly, all we need is that at least $\phi \in 2^L$).

2.1. Ordering from implications.

Nowhere did we say, up to here, that the actual world has been selected by some random process, randomness from which You could then deduce the 'probability' that the proposition ϕ is true in the actual world ω_0. Neither did we say, up to here, that you have some weighted opinions (beliefs etc....) about which world is the actual world. Without introducing such weights, all You can build if You want to quantify Your opinion is an ordering about the propositions. Suppose two propositions ϕ_1 and ϕ_2 such that ϕ_1 implies ϕ_2. In that case, whenever ϕ_1 is true so is ϕ_2, but ϕ_2 might also be true without ϕ_1 being true. So You can defend that whatever the strength

[b]Most paradoxes built against the Principle of Insufficient Reason (that would advice You to allocate a probability p to the fact that ϕ is true in ω_0) come from the arbitrariness of L. There is no absolute, definitive, 'holy' language L.

of Your opinion about the truth of ϕ_1, the strength of Your opinion about the truth of ϕ_2 must be at least as large as the one given to ϕ_1. These lead to an ordering on a lattice, not to probabilities, etc.... If You could add to this natural ordering the fact that You can compare the strength of the truth value given to any pair of propositions, so You can establish for any ϕ_1 and ϕ_2 that ϕ_1 is more, equally or less supported than ϕ_2, then the order becomes total (and it is not a big step to accept that the order could be represented by a point on $[0, 1]$).

Sugeno's fuzzy measure fits with this level of sophistication of Your opinion. Indeed, in the finite case, as studied here, Sugeno's fuzzy measure only requires that opinion be quantified on the reals, and be 'monotone for deduction', which is all You had achieved up to here.

Multi-valued logic could be used if we accept the idea that the degree of truth given to a proposition ϕ is understood as the strength You give to the fact that ϕ is true in the actual world. Furthermore, we limit ourselves to the qualitative component of multi-valued logic, the one that acknowledges the fact that some propositions are 'truer' than others and made every pair of propositions comparable for what concerns their level of truth. It is still a very shallow form of multi-valued logic as You cannot derive the truth of $\phi_1 \wedge \phi_2$ from the truth of ϕ_1 and the truth of ϕ_2 (i.e., we cannot define the strength given by You to the fact that $\phi_1 \wedge \phi_2$ is true in ω_0 from the strengths given by You to the fact that ϕ_1 is true in ω_0 and to the fact that ϕ_2 is true in ω_0). The next step is to define properties for the operators representing negation, disjunction and conjunction. If the domain of the truth value is bounded (what can also easily be assumed) then the truth functionality of these operators implies that they can be represented by triangular norms and conorms. It is essentially sufficient to assume that conjunction be associative, commutative and monotone with respect to the ordering. Adding idempotency or distributivity leads directly to the use of the minimum operator for the conjunction (and the maximum operator for the disjunction). A little extra is still needed to get the truth of the $\neg\phi$ being the complement (to 1) of the truth of ϕ.

We will not argue on the representation of the implication operator as its pragmatic meaning is not obvious. Its representation as a material implication is easily achieved but implication is not so easily captured by the material implication, especially once it is used in a modus ponens.

So we have been able to describe and present a theory dealing with imprecision. From the pure disjunctive form where we can only say 'ϕ might be true but ϕ might also be false', we have reached a level of sophistication encountered in multi-valued logic by essentially acknowledging the comparability of the truth value of any pair of propositions. The role of the triangular norms becomes primordial for developing the whole logic. But nowhere did we speak about uncertainty. We only consider imprecision with an added order on the propositions that are contingent given EC.

2.2. Ordering induced by fuzziness.

Having been able to order the worlds with regard to the truth value of ϕ, we may wonder about the origin of such ordering. The order on the lattice that results from the implication was obvious (even thought it might be so degenerated that the order is mapped on a two-value domain). The total order had to be assumed by accepting the comparability between any pair of propositions. One origin for this total order (when the domain is not degenerated to two values only) can be found in the use of 'fuzzy' propositions where fuzzy is contrasted with vague. A vague proposition is nothing but an ambiguous proposition as when I say 'I will arrive at the station at 4.15 p.m. or at 5.20 p.m.'. A fuzzy proposition does not necessarily reflect ambiguity. It requires intermediate truth values, i.e., values between true and false. When I say 'I will arrive at about quarter past four this after noon.', the statement can be true (if I arrive at 4.15 p.m.), false (if I arrive at 9.00 p.m.) but if I arrive at 4.10 p.m. the statement is more or less true, the less true, the largest the interval of time between 4.15 p.m. and the actual arrival time. Instead, vague statements can only be true or false.

ADD IMAGE HERE

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

Fuzziness is a source for 'partial truths'. A statement built with a fuzzy predicate, be it a gradable adjective or a fuzzy quantifier, is a typical example of such fuzziness. For instance, 'Paul is tall', 'John is rich', 'the speed of the car is high', 'the temperature of the furnace is low', 'few students are older than 30 years old'..... A good way to recognize a fuzzy proposition is by adding the adverb 'very' (or similar ones) to the adjective or quantifier around which the predicate is built, the resulting proposition being still meaningful. I can say 'Paul is *very* tall', 'John is *very* rich', 'the speed of the car is *very* high', 'the temperature of the furnace is *very* low', 'the arrival time is *very* close to quarter past four this after noon', '*very* few students are older than 30 years old'.... but I cannot say 'Jack is *very* dead', 'the arrival time is *very* equal to 4.15 p.m..' (dead and equal are not fuzzy). The fact that intensity adverbs like *very*, etc.... can be introduced without making the statement unacceptable is a characteristic of the fuzzy propositions.

2.2.1. Fuzzy EC and crisp ϕ.

Once fuzziness is introduced, we can then speak about an ordering of the worlds. Indeed, when building Ω_K (where K included some fuzzy propositions like 'Paul is tall' and EC includes what You understand by 'being tall') only those worlds in Ω_L where the fuzzy propositions are false, are eliminated when building Ω_K. In every world in Ω_K Paul has a particular height, there is no fuzziness within the worlds themselves. Of course, some worlds fit better than others with the requirement 'Paul is tall'. This 'fitting' results from Your understanding of 'being tall', and we assume You have only one such understanding, so we can accept that in every world in Ω_K, You have the same understanding. For every world ω in Ω_K (i.e. for every possible value of Paul's height) You can assess the degree of compatibility between the height of Paul in that world and the statement 'Paul is tall', the largest the degree, the closest Paul's height is from the height of a tall person according to Your understanding of 'tall man'. So in every world ω of Ω_K, You can assess, in theory, the compatibility of ω with EC. Let $\mu_{EC}(\omega)$ be the degree of compatibility of ω with EC where μ_{EC} maps Ω_K to $[0, 1]$ (we use the $[0, 1]$ interval for convenience only, it is arbitrary but usually perfectly acceptable). The understanding is that the largest $\mu_{EC}(\omega)$, the largest is the compatibility of ω with EC. The extreme values mean: $0 =$ not compatible at all, $1 =$ fully compatible. We speak about the compatibility of ω with EC, not with K, as the understanding of the fuzzy terms is part of Your EC without being in K.

It is straightforward to extend μ_{EC} to Ω_L by defining $\mu_{EC}(\omega) = 0$ for all ω in Ω_L but not in Ω_K. From here on the distinction between Ω_L and Ω_K is no more important as far as μ_{EC} handles the difference.

Suppose a new proposition ϕ in 2^L and let us ask if ϕ is true given EC, i.e.,what is a truth status of ϕ in those worlds in Ω_K. Each world ω is equipped with a number $\mu_{EC}(\omega)$ that represents its degree of compatibility with EC. Suppose a world ω^* in which Paul's height is h^* and let 'Paul is tall' be in EC. We understand $\mu_{EC}(\omega^*)$ as $\mu_{Tall}(h^*)$, the grade of membership of a person of height h^* to the fuzzy set of tall men according to Your understanding, and we see the whole fuzzy set theory introducing itself in our description of Your 'opinions'. Compatibility and grade of membership come to share the same scale.

Let a crisp proposition ϕ (where crisp is used in contrast to fuzzy, hence ϕ is either true or false). What might be its truth value in Ω_K given EC ? What are the equivalent of necessary and contingent when EC is fuzzy (contains fuzzy propositions). In every world ω, we know if ϕ is true or false. Let $[\phi]$ be the set of worlds in Ω_K where ϕ is true, it will not be very hard to propose that the necessary or contingent nature of ϕ given EC is a function of the value $\mu_{EC}(\omega)$ for those ω in $[\phi]$ and with some extra but natural assumptions (like those encountered when triangular norms were introduced) we get that the necessity and possibility of ϕ given EC, denoted i-N$(\phi|EC)$ and i-$\Pi(\phi|EC)$ are given by:

$$\text{i-}\Pi(\phi|EC) = \max_{\omega\in[\phi]} \mu_{EC}(\omega)$$

$$\text{i-N}(\phi|EC) = \min_{\omega\in[\neg\phi]} (1 - \mu_{EC}(\omega)) = 1 - \text{i-}\Pi(\neg\phi|EC). \tag{1}$$

When EC is not fuzzy, we could have said that ϕ is necessary if i-N$(\phi|EC) =$ i-$\Pi(\phi|EC) = 1$, and that ϕ is contingent when i- N$(\phi|EC) = 0$ and i-$\Pi(\phi|EC) = 1$.

The definition Eq. (1) has the huge advantage that it is robust and invariant to any monotone transformation of the compatibility scale. It is the only solution that satisfies monotony for deduction and that has this property that we feel must be satisfied by any 'realistic' model, where 'realistic' means related to the 'real world' in contrast to 'just mathematical' (our aim is not to create beautiful mathematical theories but theories that might be used to represent some reality, some physical, objective or subjective processes).

2.2.2. Fuzzy EC and fuzzy ϕ.

Extension to the case where ϕ is itself fuzzy is direct. Let $\mu_\phi(\omega)$ be the degree of compatibility of the world ω with Your understanding of the fuzzy proposition ϕ. So attached to every world ω in Ω_K, we have a degree of compatibility $\mu_{EC}(\omega)$ of ω with EC induced by those propositions in K and a degree of compatibility $\mu_\phi(\omega)$, both resulting from Your understanding of the various fuzzy components (those in K and in ϕ). The value $\mu_\phi(\omega)$ can also be understood as the truth of ϕ in ω, denoted $\nu_\phi(\omega)$.

Can we define i-N$(\phi|EC)$ and i-$\Pi(\phi|EC)$, the necessity and possibility of ϕ given EC? In every world ω we know $\mu_{EC}(\omega)$ and $\mu_\phi(\omega)$. It has been proposed that:

$$\text{i-}\Pi(\phi|EC) = \max_{\omega\in\Omega_K} (\mu_\phi(\omega) \wedge \mu_{EC}(\omega))$$

$$\text{i-N}(\phi|EC) = 1 - \text{i-}\Pi(\neg\phi|EC) = \min_{\omega\in\Omega_K} (\mu_\phi(\omega) \vee (1 - \mu_{EC}(\omega)). \tag{2}$$

The relations between i-N and i-Π are identical to those that link necessity and possibility in modal logic, extended to the $[0, 1]$ interval.

This possibility express the idea of the largest possible value of the truth of ϕ taken over the possible worlds under the condition that the world is compatible with EC. The use of min-max operators results again from the triangular norms requirements and the order invariance for monotone transformation of the compatibility scales.

2.2.3. Crisp EC and fuzzy ϕ.

A third case would be if there is no fuzziness in K and only ϕ is fuzzy. Then $\mu_{EC}(\omega)$ would only be 1 for all worlds in Ω_K, and 0 otherwise. So:

$$\text{i-}\Pi(\phi|EC) = \max_{\omega\in\Omega_K} \mu_\phi(\omega)$$

$$\text{i-N}(\phi|EC) = \min_{\omega\in\Omega_K} \mu_\phi(\omega)$$

Remember that $\mu_\phi(\omega)$ is the compatibility of ω with Your understanding of ϕ, and can be understood as the truth of ϕ in ω_0 The definitions of i-Π and i-N show that i- $\Pi(\phi|EC)$ and i-N$(\phi|EC)$ are the largest and the smallest values that the truth of ϕ can reach in those worlds compatible with EC.

The question is now: what do we model, imprecision or uncertainty ? In our opinion, we feel that all we have developed up to here are related to imprecision. The only delicate point is in deciding that $\mu_{EC}(\omega)$ is related to imprecision or to uncertainty. First, is it objective or subjective ? The value of $\mu_{Tall}(170)$ is the value You would personally give to the compatibility of a person of height 170 cm. with the set of tall men. Hence, it results from Your opinion about what 'tall' means. So it is subjective and personal. Is it uncertain ? For what concerns Yourself, it is not. You have an understanding of what 'tall' means. It is Yours, but You are not uncertain about it. So for what concerns You, there is no uncertainty involved. Of course, if another agent is introduced, the value You will give to $\mu_{Tall}(170)$ is uncertain for the other agent who could only express his/her belief about what will be the value You will give to $\mu_{Tall}(170)$, but this is a totally different problem. So I think, all I developed up to here is imprecision when seen under the single agent context.

There is another way to recognize that our development deals with imprecision: eliminate fuzziness from both EC and ϕ and look to what Eq. (2) becomes:

$$\text{i-}\Pi(\phi|EC) = 1,\ \text{i-N}(\phi|EC) = 1 \text{ iff } \forall\omega \in \Omega_K, \omega \models \phi$$

(ϕ is necessary in EC)

$$\text{i-}\Pi(\phi|EC) = 1,\ \text{i-N}(\phi|EC) = 0 \text{ iff } \exists\omega \in \Omega_K, \omega \models \phi \text{ and } \exists\omega \in \Omega_K, \omega \models \neg\phi$$

(ϕ is contingent in EC)

$$\text{i-}\Pi(\phi|EC) = 0,\ \text{i-N}(\phi|EC) = 0 \text{ iff } \forall\omega \in \Omega_K, \omega \models \neg\phi$$

(ϕ is impossible in EC)

So all we have compute is a criteria (the pair i-Π, i-N) that states that ϕ is contingent or not, exactly what we encountered in the case of imprecision. This limiting behavior enhances that all we have achieved by introducing fuzziness is generalizing imprecision, not introducing uncertainty.

2.3. Possibility distribution functions.

Suppose 'Paul is tall' is the only fuzzy proposition in EC. It induces the ordering we have just developed. What can we say about Paul's height ? You know for every world ω in Ω_K that $\mu_{Tall}(h) = \mu_{EC}(\omega)$ where h is the height of Paul in world ω. The possibility i-$\Pi(A)$ that h belongs to A where A is a fuzzy subset of the domain of Paul's height can be written as:

$$\text{i-}\Pi(A|Tall) = \max_{\omega\in\Omega_K}(\mu_{Tall}(h) \wedge \mu_A(h))$$

Taking A to be a singleton of the height (in which case $\mu_A(h) = 1$ if $A = \{h\}$, $\mu_A(h) = 0$ otherwise), we have i-$\Pi(\{h\}|Tall) = \mu_{Tall}(h)$, and we can define $\pi_{Tall}(h) =$ i-$\Pi(\{h\}|Tall)$. We call π_{Tall} the possibility distribution related to the possibility measure i-$\Pi(.|Tall)$, hence:

$$\text{i-}\Pi(A|Tall) = \max_{\omega\in\Omega_K}(\pi_{Tall}(h) \wedge \mu_A(h))$$

where h is the height of Paul in world ω.

Notice that the relation between i-Π and π is similar to the one between the probability distribution P and the probability density function p. Let Ω be a set and A be a crisp set. Then, in probability theory, we have:

$$P : 2^\Omega \to [0,1],\ p : \Omega \to [0,1] \text{ and } P(A) = \sum_{x\in A} p(x).$$

Identically, in possibility theory, we have:

$$\text{i-}\Pi : 2^\Omega \to [0,1], \pi : \Omega \to [0,1] \text{ and i-}\Pi(A) = \max_{x\in A} \pi(x).$$

This analogy does not mean that i-Π is automatically a measure of uncertainty, even though P is it. Again the question is: what does i-$\Pi(A)$ represent, imprecision or uncertainty? For a single agent (You), it reflects Your subjective, personal opinion about what might be Paul's height given You know 'Paul is tall' and given the understanding You give to the fuzzy predicate 'tall'. Does it quantify Your uncertainty about Paul's height ? No, it only fits with Your understanding, what is well defined for what concerns Yourself.

We had defined imprecision as a property related to the content of a statement when more than one world is compatible with the available information (and the truth status of the proposition You want to assess varies among those worlds). Imprecision is thus a property of the propositions in K and of Your understanding of them encoded in Your EC. Suppose we accept that there is nevertheless uncertainty whenever K includes a fuzzy proposition. In such a case, we must accept that imprecision is uncertainty. Indeed, if we accept that there is uncertainty when the proposition 'Paul is tall' is included in K, uncertainty quantified by i-$\Pi(A|Tall)$ for instance, then there is also uncertainty when the proposition 'Paul's height is between 160 and 170' replaces 'Paul is Tall' in K. In that last case, i- $\Pi(A|[160, 170])$, with A being a crisp subset of the reals, will be 1 if $A \cap [160, 170] \neq \emptyset$, 0 otherwise (and i-N$(A|[160, 170])$ would be 1 if $A \subseteq [160, 170]$ and 0 otherwise). This behavior is exactly the kind of statement that characterized imprecision, not uncertainty. The fact that Tall was fuzzy does not justify the shift from imprecision to uncertainty. Intrinsically, fuzziness is just a generalization of the concept of set, not a theory of weighted sets.

Of course, imprecision might be and is responsible for some further uncertainty. Once You want to express Your beliefs about which of those worlds compatible with EC is the actual world, You might not only listed the worlds in Ω_K, but you could express that some are more 'believable' than other. This would be a meta-property built atop of the imprecision that results from the fact that there are several worlds compatible with EC.

The compatibility of ω with EC, quantified by $\mu_{EC}(\omega)$, is a generalization of a characteristic function. When $\mu_{EC}(\omega)$ is seen as a function on Ω_L, and when EC is crisp, $\mu_{EC}(\omega)$ is nothing but the characteristic function of the set Ω_K as it is 1 when $\omega \in \Omega_K$, and 0 when $\omega \in \Omega_L$ and $\omega \notin \Omega_K$. A characteristic function does not represent uncertainty, only imprecision that appears when there are several ω for which the characteristic function is 1.

3. Uncertainty.

Having developed a 'full' theory of imprecision what can we do with uncertainty ? To simplify the discussion, we temporally return to crisp predicates and leave aside the fact that some predicates might be fuzzy. Uncertainty is a characteristic of the state of knowledge of the agent about which of the possible worlds is the actual world. It will be part of EC. What we introduce is an extra information that gives weights to the various subsets of Ω_K, weights that quantify the strength of Your opinion that ω_0 belongs to the various subsets of Ω_K. The overall schema obeys to the next development where we use Q to represent the strength of Your opinion about which world in Ω_K is the actual world. So for $A \subseteq \Omega_K$, $Q(\omega_0 \in A)$ can be understood as the probability, the possibility, the belief that ω_0 belongs to A according to the type of 'opinion' we consider (beware, possibility is not to be understood here as it was previously, so we denote it u-possibility in contrast to the i-possibility we used previously).

3.1. Possibility as uncertainty.

We have a set of possible worlds Ω_K and we know that one of them, ω_0, is the actual world. Which world is the actual world is not known to You ? It might occur that You have some extra ideas about which world in Ω_K is ω_0, and that these ideas reflect some 'preference' about which world in Ω_K is ω_0. This preference will be translated by saying that the fact that ω_0 belongs to A is more possible than that it belongs to B, for A and B subsets of Ω_K. If such a comparison can be performed between every pair of subsets of Ω_K, then we obtain the possibility theory where u-$\Pi(A)$ will quantify the strength You give to the possibility that $\omega_0 \in A$. Just as it was the case with imprecision, it is very easy to justify that

$$\text{u-}\Pi(A \vee B) = \max(\text{u-}\Pi(A), \text{u-}\Pi(B)) \tag{3}$$

in which case, we derive again the whole possibility theory but with another understanding. The primitive is no more in a fuzzy predicate but in an ordering about the possibilities that $\omega_0 \in A$ for every A in Ω_K. The preference ordering u-$\Pi(A) >$ u-$\Pi(B)$ translates ideas like '$\omega_0 \in A$' is more natural, normal, typical than '$\omega_0 \in B$'. The possibility measure can be seen as an ordinal measure, i.e., when only the order is relevant (like in Baconian probabilities). It is easy to justify the maximum rule Eq. (3) if the compositionality ($\Pi(A \vee B) = F(\Pi(A), \Pi(B))$), idempotency ($\Pi(A \vee A) = \Pi(A)$), associativity ($\Pi((A \vee B) \vee C) = \Pi((A \vee (B \vee C))$ and commutativity ($\Pi(A \vee B) = \Pi(B \vee A)$) are required. Troubles would appear if possibility measure is understood as a cardinal because we cannot provide a meaning to the value .7 in the statement 'the possibility is .7 that the actual world is in A'. I do not have the equivalent of a betting quotient to define what is meant by this .7 (see later on in the definition of a probability). Fuzziness should not be evoked to explain the .7, as we try to define possibility outside the scope of fuzziness. So what ? When I say, 'for all what I know, the possibility that Hans eats 3 eggs tomorrow morning for breakfast is .8', what is this .8 representing ?

Suppose we speak again about the weight of Paul and we accept cardinal u-possibilities. Why the .7 in 'i-$\Pi([150, 170]|Tall) = .7$' is an imprecision whereas in 'u-$\Pi([150, 170]|[160, 180]) = .7$' it quantifies uncertainty. Why i-$\Pi([150, 170]|Tall)$ is not a measure of uncertainty. We said that we had to see what occurs to the possibility when fuzziness is taken away. So suppose that Tall means [160,180] for You. Then i-$\Pi([150, 170]|[160, 180])$ will not be .7 but 1. In fact, i-$\Pi(A|B) \in \{0, 1\}$ if both A and B are crisp. Take away fuzziness and intermediate values for i-Π disappear, which is not the case for u-Π . This also enhances the fact that u-Π is derived from extra assumptions present in EC : the preference ordering.

For the ordinal u-Π, the distinction is not so clear. Indeed, ordinal u-Π only creates some order on Ω_K. But we saw that such an order could also be achieved if we only accept monotonicity for deduction and comparability. In fact, we had introduced that ordering in the imprecision section and explains its origin from the presence of some fuzziness. Such an ordering must be satisfied by any representation of imprecision and of uncertainty, once we accept a domain that tolerates intermediate values. In the imprecision section, the ordering on $[0, 1]$ instead of $\{0, 1\}$ collapses once fuzziness was eliminated, whereas in the cardinal u-Π, the ordering on $[0, 1]$ persists in the absence of fuzziness. So in both cases, cardinal and ordinal, i-possibility and u-possibility are not identical concepts.

We could still try to claim that u-Π and i-Π are not different concepts by further degenerating the u-Π such that there is a unique set $A^* \subseteq \Omega_K$ with for all A in Ω_K, u-$\Pi(A) = 1$ if $A \cap A^* \neq \emptyset$, and u-$\Pi(A) = 0$ if $A \cap A^* = \emptyset$. We could end up with a partition on Ω_L made of A^* , Ω_K/A^* and Ω_L/Ω_K, three sets that are usually non empty.

With i-Π , the set A^* is equal to Ω_K by definition. Indeed, with imprecision, which we got out of EC was the set Ω_K, and for all subset A of Ω_K, we deduced i-$\Pi(A) = 1$ if $A \cap \Omega_K \neq \emptyset$ and 0 otherwise. Instead with uncertainty, EC produces not only Ω_K but also the set A^* , and we deduce that u-$\Pi(A) = 1$ if $A \cap A^* \neq \emptyset$, and 0 otherwise.

In this highly degenerated form of u-possibility, we face a set A^* of worlds in Ω_K preferred to those in Ω_K/A^* . The problem is to give a meaning to the worlds in Ω_K/A^* in the u-possibility context. We know the worlds in A^* are possible (given EC), those in Ω_L/Ω_K are impossible (given EC). Those in Ω_K/A^* are compatible with K, but are considered as admitting a zero possibility given EC. Their 'impossibility' does not reflect a property in K but in EC/K, whereas for those worlds in Ω_L/Ω_K their impossibility (in both i- and u-possibility cases) reflects a property in K.

The distinction might become clear if You revise EC by adding the information $\omega_0 \notin A^*$ to EC (and therefore to K). The u-possibility would have to be revised in order to cope with this new information, and You would admit as possible now that $\omega_0 \in \Omega_K/A^*$. In the i- possibility case, if You add the information $\omega_0 \notin A^*$ to EC, there is no way to save EC from inconsistency (except by some drastic change in EC that consists in eliminating some pieces of information in order to restore consistency).

We could state that u-possibility is an 'epistemic' construct as u-possibility pertains to EC/K whereas the i-possibility is a 'logical' construct as it pertains directly to K, except for the fact that the logical deduction on 2^L obtained from K were themselves based on the epistemic construct EC. The real difference is that in the highly degenerated case, the i-possibility depends on EC through K whereas the u-possibility depends on EC through K and EC/K.

Probably the u-possibility and the i-possibility can be assimilated to the epistemic and the physical possibilities that were described by Hacking.

3.2. Probability.

3.2.1. Objective Probability.

The next theory about uncertainty can be constructed if we accept that the actual world ω_0 has been selected in Ω_K by a chance set up. If ω_0 corresponds to the face of a coin that comes up when tossed, then You include this information in EC. You can allocate to every subset A of Ω_K a weight that quantifies the 'chance' that $\omega_0 \in A$. These weights are probabilities and satisfied Kolmogorof axioms for probability functions. Are these probabilities objective or subjective? As described so far, they look objective and exist outside of You. If we accept the concept of objective probabilities, we obtain a theory of objective uncertainty which properties are well established, but which meaning is seriously questioned by some philosophers. What is an objective probability? Is it true that chance really exists, without involving an agent? Are there random events in physics that exists without introducing the observer's uncertainty or limited knowledge about reality? Is the world deterministic? These questions are still unsettled.

3.2.2. Subjective probabilities.

Some Bayesian extremists will claim that objective probabilities do not exist. They claim that the probability .5 given to heads when tossing coins is neither a property of the coin nor of the experimental set up that generates the outcome (my thumb, the coin, me tossing the coin, the direction of the wind, the rotation of the Earth....). The .5 is a property of the agent, You, who observes the experience. That the coin falls heads or tails is a deterministic outcome, but due to Your limited understanding, You cannot predict exactly the outcome and You can only express Your subjective, personal opinions about the outcome. Then why .5 ? What does it mean ?

3.2.3. Betting quotients.

The answer is to be found in the betting quotients and a study of Your behavior when You must decide. The probability p of an event has been defined as a price You would pay to enter a game where You would gain one unit of money if the event occurs and nothing otherwise. Besides the bets must be so organized that the probability assessor gives the value p and his opponent decides if the probability assessor will be the player or the banker (the player pays p and gets one unit of money if the event occurs, the banker receives p and pays one unit of money if the event occurs).

Refusing the bets is not allowed. A close analysis of the consequences of the forced bets scheme shows that the only way the probability assessor can avoid a Dutch Book (a set of bets that leads necessary to a sure loss whatever event occurs) is by assessing betting quotients that satisfies probability functions assumptions. The idea of betting quotients gives a meaning to the value of the probability and the Dutch Book argument explains why betting quotients must be probability functions.

Equivalent results are not yet well established for possibility theory but the lack of results cannot be used to reject that theory. Indeed probability theory was invented by Pascal but it took a century to Bayes to propose the definition of a probability as a betting quotient and in practice the betting quotient and Dutch Book argument were developed in the 20's century. It took two hundreds years to justify probability theory. We can excuse the developers of possibility theory for not having produced their own justification within the 15 years since their model was imagined.

Once subjective probability is acknowledged, the idea of a chance set up is not necessary. You know ω_0 is in Ω_K and You might be ready to bet on $\omega_0 \in A$ for every A in Ω_K. In that case, You would assess Your betting quotients, i.e., the probabilities that the actual world belongs to A for every A in Ω_K (using probability measures to avoid to become a money pump).

This measure surely covers the domain of uncertainty. You have added extra information about which world is ω_0 and the information is represented by a probability measure. But does this model cover every forms of uncertainty? Yes claim many Bayesians. I feel the answer is not so definitive and I want to present alternative models.

3.2.4. Probability of a fuzzy event.

The study of uncertainty measure can be extended to the case of fuzzy proposition or propositions which truth value is in $[0,1]$ and not restricted to $\{0,1\}$ as in classical logic. Let $\nu(\phi|\omega)$ be the true value of ϕ in world ω: $\nu : 2^L \times \Omega_K \to [0,1]$. When Ω_K is such that all You know is that $\omega_0 \in \Omega_K$, then all You know for what concerns the truth of ϕ is that $\nu(\phi|\omega_0) \in \{\nu(\phi|\omega) : \omega \in \Omega_K\}$ and usually $\nu(\phi|\omega)$ will vary among the world of Ω_K in which case $\nu(\phi|\omega_0)$ is imprecise. There would be no imprecision in ϕ if $\nu(\phi|\omega)$ is constant for all $\omega \in \Omega_K$, which against shows that the fuzziness of ϕ is not in itself responsible for the imprecision. As already mentioned, fuzzy sets theory is just an extension of sets theory and as such, it is not imprecision, exactly as sets theory is not a theory of imprecision.

Now suppose, we have a probability measure over Ω_K (we analyze the probability case as being the simplest and most common). When the truth domain of ϕ is $\{0,1\}$ (denoted $\{T,F\}$) , we can choose among the following expression that will translate the same ideas. Let ϕ and ψ be two propositions in 2^L.

$$P((\nu(\phi|\omega_0) = T) \wedge (\nu(\psi|\omega_0) = F)) = .65$$

or equivalently:

$$P(\omega_0 \in [\phi] \cap [\neg\psi]) = .65.$$

We could also write:

$$P(\phi \wedge \neg\psi) = .65.$$

Suppose the truth domain is $[0, 1]$, and let $\tau_1 \subseteq [0, 1]$ and $\tau_2 \subseteq [0, 1]$ be two subsets of the truth domain. We get:

$$P((\nu(\phi|\omega_0) \in \tau_1) \wedge (\nu(\psi|\omega_0) \in \tau_2)) = .65 \tag{4}$$

or equivalently:

$$P(\omega_0 \in [\phi]_{\tau_1} \cap [\psi]_{\tau_2}) = .65$$

where $[\phi]_{\tau_1}$ is the set of world in Ω_K such that $\nu(\phi|\omega) \in \tau_1 \subseteq [0, 1]$ (and similarly for $[\psi]_{\tau_2}$)

We cannot write anymore the equivalent of $P(\phi \wedge \neg\psi)$ as the negation of '$\nu(\phi|\omega) = .8$' is not '$\nu(\phi|\omega) = .2$' but '$\nu(\phi|\omega) \neq .8$' (where truth domain is $\{0, 1\}$, 'not 1' is '0' but when the domain is $[0, 1]$, 'not .8' is 'anything but .8'.) This remark is obvious but its oversight has been at the origin of some nice errors.

The importance of this representation of P is that we end up with the probability that deals with the localization of ω_0. The .65 in Eq. (4) tells that the probability is .65 that ω_0 belongs to a subset of worlds in Ω_K. Identical relations are obtained with belief functions and also with u-possibility functions in which cases we end up with $bel(\omega_0 \in A) = .65$ or u-$\Pi(\omega_0 \in A) = .65$, where the .65 quantifies the belief or the possibility that ω_0 belongs to a subset of worlds in Ω_K. Using a crisp proposition ϕ does not change the scheme. The uncertainty is still there. This contrasts with what we had encountered with i-possibility theory when we spoke about imprecision (probability and belief are never confused with imprecision and the only real problem is with possibility theory because confusion can be immediate).

When we spoke about i-possibility theory in the imprecision context, we had defined i-Π with a crisp EC and a fuzzy ϕ, as:

$$\text{i-}\Pi(\phi|EC) = \max_{\omega \in \Omega_K} \nu(\phi|\omega)$$

It is not a measure of a subset of Ω_K, it is just the upper limit (with i-N being the lower limit) of the possible values of $\nu(\phi|\omega)$ in those worlds ω in Ω_K. It does not quantify the fact that ω_0 belongs to some subsets. It just says: whatever ω_0, $\nu(\phi|\omega)$ will be at most i-$\Pi(\phi|EC)$ and at least i-N$(\phi|EC)$. All i-Π and i-N achieve is a characterization of the imprecision we have about the value of $\nu(\phi|\omega_0)$, not about which world in Ω_K is ω_0.

3.3. Non-standard probability models.

We first consider the non-standard probability models, i.e., models that accept an extended concept of probability functions. One of the assumptions defended by the Bayesians is that for every event whatsoever, You can define - in theory - its probability, in which case the only remaining problems are to assess these probabilities and to apply correctly the rule of probability theory.

3.3.1. Upper and lower probabilities.

Some authors defend the idea that probabilities are not so well known. There are cases, so they claim, where some objective probabilities exist but it is not known exactly by You. All You know, is that the probability is between some boundaries. This reflects nothing but imprecision about the value of some parameters (where the parameter is the objective probability of the event under consideration). All we said about imprecision can be introduced here and we derive a theory of imprecise probability. In most practical cases, it turns out to be a theory of upper and lower probabilities.

3.3.2. Meta-probability.

We could add some uncertainty measure about the value of the parameters and as the parameter is a probability, the uncertainty measure will be called a meta-probability or a second order probability. In that case, we are back to the classical probability model and everything is in theory well defined except for a serious philosophical problem : what is a betting quotient when the event is the value of a probability ? We have a first event A at the object level, we want to assess P(A), the objective probability of A that we don't know exactly. We say that the meta probability that '$P(A) \in [a, b]$ is .7' where .7 is supposed to be the betting quotient. This is perfect..... if someone can decide if I win or if I loose. So there must be a procedure capable of deciding if $P(A) \in [a, b]$? and the procedure should be objective. How is it going to be established ? How to decide that it is true or not ? The question and the lack of appropriate answer explain the discomfort encountered with regard to meta-probability theory even though the model is mathematically elegant, beautiful and convenient.

3.3.3. Dempster's model.

A special form of upper and lower probability was developed by Dempster who examined the probability induced by a one-to-many mapping. Mathematically, Dempster's model is essentially a model based on random sets. Its importance today comes from the so-called Dempster-Shafer's model defended by Shafer (not in his book but in some of his later papers) where he claims that beliefs on a frame of discernment Ω are quantified by belief functions themselves generated by an underlying probability on some space X and a compatibility relation (a one-to-many mapping) between the space X and the frame of discernment Ω. The hints model of Kohlas and Monney fits with Dempsters approach.

3.3.4. Family of probability measures.

Still other authors suggest that the belief's state of an agent is not represented by a probability measure but by a family of probability measures. The difference with the upper and lower probability as seen before is that in the previous case, we had assumed the existence of a probability measure and the ignorance by the agent on the exact value of the probability measure. In the family of probability measure model, there is no particular but unknown underlying probability measure. There is just a family of probability measures that characterizes the agent's belief.

3.4. Non-probabilistic models: the transferable belief model.

Finally, we have other models for uncertainty not based on probability functions and classified as non-probabilistic models. We already saw the model based on possibility functions. We now examine the transferable belief model, a model for representing quantified beliefs built without requiring any underlying hidden probability measures. The transferable belief model is essentially what Shafer's book present, except Shafer insists more on the sources that induce beliefs. In fact the two approaches are so strongly linked that they can be equated.

When justifying the subjective probability, we insisted on the fact that the bets were forced bets. Besides, we had also to insist on the fact that the agent was ready not only to bet on any event, on the truth of any proposition, but further more that beliefs outside such betting contexts did not exists (when betting has to be taken in a very general sense). A contrario, I claim that belief can be held, can be entertained without having to decide. But if such state of belief exists, as I defend it, then the betting quotient and the Dutch Book arguments do not apply anymore. Hence probability measure looses its justification. All we know is that once decision will have to be made, we will have to build a probability measure on the possible alternatives in order to make coherent, rational decisions (i.e., to avoid Dutch Books and similar problems).

In the transferable belief model, I have introduced the distinction between a credal level, i.e., a level where beliefs are only entertained, and a pignistic level, i.e., where beliefs are used to make decision. At the pignistic level, uncertainty must be represented by a probability measure if we accept the Bayesian justification (personally I accept it). At the credal level, nothing justifies the use of probability measure. All that will be needed is a strategy to build the appropriate probability measure to make decision when decision will have to be made. The transferable belief model is exactly such a model where I assess that beliefs at the credal level is represented by a belief function and I have developed and justified the transformation that applied to a belief function generates the probability measure needed for decision-making. The axiomatic justification of the use of belief functions at the credal level instead of any other measure is based essentially on the idea that the belief state of an agent over the frame of discernment is fully characterized by the belief function over Ω_K and that the way the belief was reached was irrelevant when beliefs must be revised for new knowledge (the conditioning process).

The transferable belief model is thus a model that expresses the agent's belief that $\omega_0 \in A$ for every A in Ω_K. Its extension to other non- epistemic problems is immediate, just like the extension of subjective probability to classical, physical processes. So one should not restrict the transferable belief model only to subjective problems. One can implement it in any reasoning structure, like a robot or a computer, in order to take optimal decision in context where randomness and objective probability are irrelevant. A major property that we observe in the transferable belief model is that it degenerates smoothly and nicely into the classical Bayesian model whenever all the ingredients needed by the Bayesian models become present. The advantage of the transferable belief model resides in its ability to handle any level of uncertainty up to the level of total ignorance whereas the Bayesian model requires a lot of knowledge (like the knowledge of the probability of every event).

Bayesian and other probabilistic models often require an artificial probabilization, i.e., the unjustified assumption of the existence of some probabilities.

4. Conditioning.

A model for the representation of quantified beliefs can not be completed if we do not study how it behaves when new pieces of information become known to You. Just to show the impact of such dynamic component, we consider three models for uncertainty, and how they react when Your EC is revised by an information that says that the actual world ω_0 does not belong for sure to some subset $[\neg\xi]$ of Ω_K. This information corresponds to the conditioning process on ξ.

4.1. Static component.

We now consider how probabilities and beliefs can be defined in a possible worlds context. Let Ω_K be a set of world, one of them being the actual world ω_0. In every world, every proposition is either true or false. Suppose a proposition ϕ and let $[\phi]$ be the set of worlds in Ω_K where ϕ is true: $[\phi] = \{\omega : \omega \in \Omega_K, \omega \models \phi\}$. We write

$\phi \models \psi$ if $\omega \models \psi$ for every ω in Ω_K such that $\omega \models \phi$.

4.1.1. Probability functions.

Suppose we put non negative weights $p(\omega)$ on each worlds $\omega \in \Omega_K$ such that the weights add to one. These weights $p(\omega)$ are the values of a probability distribution function on Ω_K. The probability $P(\omega_0 \in [\psi])$ that the actual world ω_0 belongs to $[\psi]$ for $[\psi]$ subset of Ω_K (or equivalently the probability that ψ is true in the actual world ω_0) is obtained by adding the weights given to the worlds ω of Ω_K where $\omega \models \psi$. Such a definition of the probabilities satisfies Kolmogoroff axioms (P is additive and $P(\Omega_K) = 1$).

4.1.2. Transferable belief model.

Suppose we put non negative weights $m([\phi])$ on each subset $[\phi]$ of worlds in Ω_K (representing the proposition ϕ) such that the weights add to one. These weights $m([\phi])$ are the values of a basic belief assignment defined on Ω_K. The weight $m(A)$ represents the part of belief that supports that ω_0 belongs to A, does not support that ω_0 belongs to B where B is any strict subset of A, but that could support B if further information would justify it. Suppose a proposition ψ. The belief $bel(\omega_0 \in [\psi])$ that the actual world ω_0 belongs to $[\psi]$ for $[\psi] \subseteq \Omega_K$ (or equivalently the belief that ψ is true in the actual world ω_0) is obtained by adding the weights given to the subsets of worlds $[\phi]$ of Ω_K for those ϕ such that $\phi \models \psi$ and $\phi \neq \perp$. Given such a definition, bel is a belief function.

4.1.3. Probability of believing.

Suppose we study a set of possible worlds Ω_K and a modal operator $\Box$ that means 'I believe'. Let R be the accessibility relation described in Kripke semantics, and let $R(\omega)$ denotes the set of worlds in Ω_K accessible from ω: $R(\omega) = \{\omega' : \omega' \in \Omega_K$ such that $\omega R \omega'\}$. As in the probability case, each world ω in Ω_K is equipped with a non negative weight $p(\omega)$ and the $p(\omega)$'s add to 1. The probability $P([\Box\psi])$ that You believe ψ (or more specifically, the probability that You believe that ψ is true in the actual world) is defined as the sum of the weights given to the worlds of Ω_K where 'You believe ψ' holds. If we write $Q([\psi]) = P([\Box\psi])$, then it can be shown that Q is a belief function defined on Ω_K. The $\Box$ operator could as well have been defined as 'You know', 'You prove', 'is necessary', without changing the results.

4.2. Dynamic component.

We now study the impact of a conditioning event. It will enhance the difference between the transferable belief model and the probability of believing solutions.

Suppose You learn for sure that ω_0 is not in $[\neg\xi]$, (what would mean that ξ is true in the actual world under the condition that there is at least one world in Ω_K where ξ was true, i.e., $[\xi] \cap \Omega_K \neq \emptyset$).

4.2.1. Probability Theory.

The conditioning information implies that the probability given to $[\neg\xi]$ should become zero, which is achieved by putting all $p(\omega)$ to zero for those ω where ξ is false, and rescaling the remaining probabilities in order to keep the probabilities normalized. The requirement $P([\xi]) > 0$ translates essentially the fact that $[\xi] \cap \Omega_K \neq \emptyset$. The case $P([\xi]) = 0$ is usually left aside. The result is the classical Bayes rule of conditioning where the probability that the actual world is in $[\psi]$ given it is in $[\xi]$, is given by:

$$P([\psi] \mid [\xi]) = \frac{P([\psi] \cap [\xi])}{P([\xi])}$$

4.2.2. Transferable belief model.

Suppose a proposition ψ. The conditioning information implies that the actual world does not belong to those worlds in $[\psi]$ where ξ is false. Hence the weight that was given to $[\psi]$ is transferred to $[\psi \wedge \xi] = [\psi] \cap [\xi]$. Such a transfer leads to the so called Dempster's rule of conditioning described in Dempster-Shafer theory (where normalization is applied but $bel([\neg\xi]) < 1$ is required) and in the transferable belief model (where no normalization is applied, hence conditioning on ξ when $bel([\neg\xi]) = 1$ is manageable).

Unnormalized conditioning is given by:

$$bel([\psi] \mid [\xi]) = bel([\psi] \cup [\neg\xi]) - bel([\neg\xi])$$

and its normalized version is given by:

$$bel([\psi] \mid [\xi]) = \frac{bel([\psi] \cup [\neg\xi]) - bel([\neg\xi])}{1 - bel([\neg\xi])}.$$

4.2.3. Probability of believing.

The impact of the conditioning event is not so obvious as in the previous case. We could consider the hypothetical conditioning where we want to assess the probability that You would believe that the actual world is in ψ if You restrict Yourself among those worlds where You believed ξ. The result would be the geometrical rule of conditioning, i.e.,

$$bel([\psi] \mid [\xi]) = \frac{bel([\psi] \cap [\xi])}{bel([\xi])}.$$

Another case would be to compute the probability that You believe that the actual world is in $[\psi]$ after You add in every world that the actual world is in $[\xi]$, hence after changing the accessibility relation such that only those worlds where x is true are accessible. So the accessibility relation R is transformed into the new accessibility relation R_ξ that satisfies $R_\xi(\omega) = R(\omega) \cap [\xi]$. Then the probability that the actual world belongs to $[\psi]$ is the sum of the probabilities given to those worlds of $\Omega_K \cap [\xi]$ where $\Box\psi$ holds according to the new accessibility relation R_ξ. The results is again Dempster's rule of conditioning.

What happens with those worlds ω where $R(\omega) \cap [\xi] = \emptyset$? Suppose a set of worlds D such that $R(\omega) \cap [\xi] = \emptyset$ for every ω in D and with weight $m(D) > 0$. The weight $m(D)$ initially given to D is now given to a set of worlds that are in relation through R_ξ to no world, so every proposition (even $\perp$) is 'necessary' and none is 'possible' in those worlds in D. The problem is to know how to define $\Box\psi$ when we want to express 'I believe ψ'. If we want that 'I believe ψ' implies that 'I don't believe $\neg\psi$', then the weight $m(D)$ given to D should not be included in the probability that I believe ψ once I learn ϕ. Should we keep them and define a weight given to the contradiction (what is indeed believed in those worlds in D) as I defend it in the transferable belief model, or should we 'throw them away' and renormalized as done by Shafer is still an debatable question.

In any case, as can be seen, the impact of the conditioning is not always obvious. In probability theory and in the transferable belief model, the solution is quite straightforward, in the probability of believing, things become less obvious. The last case justifies why I insist on the fact that theories should not be compared only on their static components, but that their dynamic components must also be examined before assimilating them. This criticism is essentially oriented toward those who assimilated Dempster-Shafer theory and the transferable belief model to a theory of random sets, to a theory of upper and lower probability, to a theory of probability of provability. There are subtleties that must be correctly understood, and many appear only once conditioning is analyzed.

5. Conclusions.

After this overview of the many models for imprecision and uncertainty, we should stress that we have only tackled with their static structure, i.e.,how to represent the state of knowledge, of belief, of opinion of an agent at a given time and with the conditioning, a particular form of dynamic. Other dynamic component can be considered. Conditioning as considered classically in probability theory is just one very particular form of dynamic. Revision, focusing, updating are more general concepts than conditioning. We did not cover these problems here but we insist on the fact that often models look similar at the static level (leading authors to equate them) even though they completely diverge once their dynamic is analyzed. A real comparison between models must be performed at least at both the static and the dynamic level in order to be comprehensive.

6. Acknowledgements

This work has been partly supported by the CE-ESPRIT III Basic Research Project 6156 (DRUMS II), and the Communauté Française de Belgique, ARC 92/97-160 (BELON).