TR－A－0112

Mathematical Connections between the probability, Fuzzy set, Possibility and Dempster-Shafer theories

Philippe QUINIO

1991. 4. 12

ATR視聴覚機構研究所

〒619-02 京都府相楽郡精華町乾谷 ☎07749-5-1411

ATR Auditory and Visual Perception Research Laboratories

Inuidani,Sanpeidani, Seika-cho, Soraku-gun, Kyoto 619-02 Japan

Telephone: +81-7749-5-1411
Facsimile: +81-7749-5-1408
Telex: 5452-516 ATR J

Mathematical Connections between
the probability, Fuzzy set, Possibility and Dempster-Shafer theories

Contents

Mathematical Connections between

the probability, Fuzzy set, Possibility and Dempster-Shafer theories

Philippe QUINIO

Auditory and Visual Perception Research Laboratories

Advanced Telecommunications Research Institute

1. Introduction and Outline

Expert systems have been regarded by many as the single most important contribution of Artificial Intelligence (AI) to engineering and science in general. The AI community however, has been increasingly aware, over the past few years, of the fundamental limitations of rule-based systems when trying to match human beings in solving complex problems. It has also become increasingly clear that these limitations are at least partly related to the inability of standard (predicate) logics to deal with uncertain, imprecise or fuzzy information.

The need for a new framework that would overcome these difficulties has triggered a huge amount of research since the pioneering works of L.Zadeh (1965) and G.Shafer (1976, [1]). The result of these efforts was the proliferation of "new theories" for the representation of uncertainty, imprecision or fuzziness in AI, including the Upper/lower probability framework, Dempster-Shafer (DS) theory of Evidence (Belief/Plausibility functions), Possibility theory (Possibility/Necessity measures), Fuzzy set and Random Closed Set (RACS) theories...

The authors of these theories spent much of their energy trying to isolate them and prevent external criticism or comparison with other scientific theories. This is quite natural under such circumstances and, as a matter of fact, played a positive role as a protective shield for their growth in the early stages.

We believe however that the time has come for a systematic in-depth theoretical comparison. The mathematical links between the Random Closed Set formalism, of which section 2 recalls the basic conceptual and technical background, and topological versions of the above mentioned theories are investigated in detail in section 3, the basis for comparison being purely axiomatic. The RACS theory emerges as a *sufficient* conceptual and mathematical framework for the representation of uncertainty, imprecision and fuzziness. The underlying topological setting makes it sufficiently general so as to encompass all AI problems, but not so general so as to include useless, experimentally inaccessible mathematical abstractions.

Section 4 provides a discussion about the necessity and the usefulness of theoretical comparisons in the context of Artificial Intelligence, with a particular emphasis on the axiomatics/interpretation dilemma.

Since the RACS theory is merely an application of general probability measure theory to a topologically meaningful subpart of the power set of a Universe $\mathcal{U}$ equipped with a topology derived from that of $\mathcal{U}$ itself, the tone of this paper might appear somewhat retrograde. We insist, however, on the practical usefulness of the above theoretical comparisons, that go far beyond mere scientific conscientiousness. And we show at length in section 5 how such comparisons can be exploited in order to compensate for theoretical weaknesses in the above formalisms and yield useful hybrid techniques. An important result characterizes the Mean probabilistic operation as the only order-independent, piecewise and point-compatible combination for constructing general Belief, Plausibility or Fuzzy set membership functions from subsets of a Universe. This reduces the possible choices of a construction scheme in systems where the sources of information are not human, or at least not *only* human.

2. Conceptual and technical background in RACS theory

2.1 Imprecision, Uncertainty and Random Sets

Imprecision is a *set-theoretic concept*: a piece of information is said to be *imprecise* (with respect to a given Universe $\mathcal{U}$) if it can be represented by a **subset** of $\mathcal{U}$ but not by a single element of $\mathcal{U}$. In physics, for instance, ($\mathcal{U} = \Re$), a measurement is said to be imprecise whenever its value lies within an "error interval", and in fact, the result of the measurement is not a real value but the interval itself.

Uncertainty is a *probabilistic concept*, and in fact, the whole theory of probability measures was precisely created to quantify uncertainty. Several recent works ([2], [3]) suggest that the axioms of probability theory are in fact the only ones that are compatible with the usual intuitive concept of "uncertainty". The measure-theoretic concept of measurable mapping leads to the fundamental tool of *random variable* from a measure space to a measurable space. When working in a Universe $\mathcal{U}$, the straightforward (and classical) approach is to use random variables taking values in $\mathcal{U}$ itself, i.e. **random points** of $\mathcal{U}$.

When dealing with problems that involve both imprecision *and* uncertainty, it seems natural to use **random sets** instead of random points of $\mathcal{U}$. A random set is a measurable mapping from a probability space $(\Omega, \Sigma_\Omega, \mathbf{Prob})$ into a measurable space $(\mathcal{P}(\mathcal{U}), \Sigma_{\mathcal{P}(\mathcal{U})})$ where $\Sigma_{\mathcal{P}(\mathcal{U})}$ is a σ-algebra over $\mathcal{P}(\mathcal{U})$. Like any other random variable, a random set X is entirely determined by its probability distribution P' defined on $\Sigma_{\mathcal{P}(\mathcal{U})}$:

$$\forall A \in \Sigma_{\mathcal{P}(\mathcal{U})}, \qquad P'(A) = \mathbf{Prob}(X^{-1}(A)) = \mathbf{Prob}(\{\omega \in \Omega,\ X(\omega) \in A\}) \qquad \text{noted} \quad \mathbf{Prob}(X \in A)$$

For instance, [4] (p.41) considers the σ-algebra generated by the sets $M_I = \{A \in \mathcal{P}(\mathcal{U}),\ I \subset A\}$ and $M^{I'} = \{A \in \mathcal{P}(\mathcal{U}),\ I' \cap A = \emptyset\}$ where $I, I' \in \mathcal{I}$ are two *finite* subsets of $\mathcal{U}$. From the classical Kolmogorov Extension theorem, it follows that any random set X defined with this σ-algebra is entirely determined by its *space law* T': $T'(I) = \mathbf{Prob}(X \cap I \neq \emptyset) \quad \forall I \in \mathcal{I}$ (see also [5]).

We can also take $\Sigma_{\mathcal{P}(\mathcal{U})} = \mathcal{P}(\mathcal{P}(\mathcal{U}))$, in which case we obtain the most general class of random sets.

2.2 Topology

We believe, however, that Set theory is too general for practical problems, and that *topology is needed* to select the only physically plausible (and practically useful) sets from the huge number of general subsets of a (possibly uncountably infinite) Universe. As a matter of fact, no experiment will ever be able to distinguish between the real and the rational numbers, for example, and thus, for all practical purposes, we might just as well merge these two concepts. More generally, we might as well merge any set $A \subset \Re$ with its topological closure $\bar{A}$, and its topological interior $\dot{A}$ (the topology considered here being induced by the usual Euclidean metric in $\Re$).

There is yet another reason why we need topology: we would like to give a mathematical meaning to the intuitive concept of the *continuity* of a mapping and this is done by introducing a topological structure on both the definition and the value sets of the mapping. Similarly, the intuitive concept of the *convergence* of a sequence of points is defined rigorously in a topological framework.

Thus, we shall assume that $(\mathcal{U}, \mathcal{T})$ is a *topological space*, $\mathcal{T}$ being the set of opens of $\mathcal{U}$. In general, one can define many different topologies on the same Universe $\mathcal{U}$, most of them being trivial or "pathological cases" (such as a non-discrete topology in a finite set, or the coarse topology where only constant functions are continuous...). As these topologies do not suit our purposes (what would be the use of a topology that makes all non-constant functions discontinuous?!), we have to discard them by assuming a few basic properties on $(\mathcal{U}, \mathcal{T})$: the Hausdorff property and local compactness. Then, we explicitly state that *two subsets with the same closure and the same interior are undistinguishable*, which leads to considering a quotient space modulo this equivalence relation. This quotient space may be identified with a subset of $\mathcal{O}(\mathcal{U}) \times \mathcal{F}(\mathcal{U})$, which naturally leads to either the theory of Random *Closed* Sets

(RACS) or the theory of Random *Open* Sets (RAOS). Both are basically equivalent to each other and we may arbitrarily select RACS.

2.3 Random Closed Sets and the Choquet theorem

Starting from a **locally compact Hausdorff**[1] topological space $(\mathcal{U},\mathcal{T})$, we can equip the set $\mathcal{F}(\mathcal{U}) = \mathcal{U}'$ of all the **closed** subsets of $(\mathcal{U},\mathcal{T})$ with a topological structure $\mathcal{T}'$, called the **Hit or Miss topology**: it is the topology generated by the (open) sets $O'_O = \{F \in \mathcal{F}(\mathcal{U}),\ F \text{ hits } O\} = \{F \in \mathcal{F}(\mathcal{U}),\ F \cap O \neq \emptyset\}$ for all opens $O \in \mathcal{T}$ and by the (open) sets $O'^K = \{F \in \mathcal{F}(\mathcal{U}),\ F \text{ misses } K\} = \{F \in \mathcal{F}(\mathcal{U}),\ F \cap K = \emptyset\}$ for all compacts $K \in \mathcal{K}$. The opens $O' = O'_{O_1} \cap O'_{O_2} \cap \cdots \cap O'_{O_n} \cap O'^K = \{F \in \mathcal{F}(\mathcal{U}),\ F$ **hits** $O_1, \ldots, O_n$ and **misses** $K\}$ where $O_1, \ldots, O_n$ is a finite family of *opens* and K is a *compact* of $\mathcal{U}$, form an open base for $\mathcal{T}'$.

The intuitive idea behind this topology is as follows: the more two closed sets F_1 and F_2 **hit** the given family of opens $O_1, O_2, \ldots, O_n$ and **miss** (i.e. fail to hit) a given compact K, the more they are said to be "neighbors" (figure 2). It can be shown ([4], p.3) that $\mathcal{U}'$ equipped with the Hit or Miss topology $\mathcal{T}'$ above is **compact** (and Hausdorff), which guarantees the existence of topological probability measures on $\mathcal{U}'$.

A **Random Closed Set** (RACS) of $(\mathcal{U},\mathcal{T})$ is simply a random variable defined on an underlying probability space $(\Omega, \Sigma_\Omega, \mathbf{Prob})$ and taking values in the (compact) measurable topological space $(\mathcal{F}(\mathcal{U}), \Sigma')$, where Σ' is the Borel σ-algebra of $(\mathcal{U}', \mathcal{T}')$. The **distribution** P' of a RACS X is such that: $\mathbf{Prob}(X \in A') = P'(A')$ for every event A' in the σ-algebra Σ' and entirely determines X.

The above definitions should be sufficient to elaborate a formal theory of Random Closed Sets. However, such a theory would not go very far unless an additional property of $(\mathcal{U},\mathcal{T})$ is assumed. As a matter of fact, we shall be very much concerned with *continuity*: a mapping $f : \mathcal{F}(\mathcal{U}) \longrightarrow \Re$, for example, is continuous iff the inverse images $f^{-1}(O)$ of any open set $O \subset \Re$ is an open set in $\mathcal{F}(\mathcal{U})$. This can be simplified *if the topology $\mathcal{T}'$ of $\mathcal{F}(\mathcal{U})$ admits a countable base* of open sets: the real sequence $\{f(F_n)\}$ converges in $\Re$ towards $f(F)$ whenever a sequence $\{F_n\}$ converges in $\mathcal{F}(\mathcal{U})$ towards a limit F. It can be shown ([4] p.3) that if $(\mathcal{U},\mathcal{T})$ is **second countable** (i.e. $\mathcal{T}$ admits a countable base) and locally compact, then $(\mathcal{F}(\mathcal{U}), \mathcal{T}')$ is also second countable (and locally compact). By a well-known result, this implies that both spaces are *metrizable*, i.e. there exists for each of them a metric compatible with its respective topology (see figure 1 and [6]). This remark greatly simplifies the convergence criterion and hence most continuity problems: from the definition of the Hit or Miss topology, we know that a *sequence $\{F_n\}$ converges in $\mathcal{F}(\mathcal{U})$ towards a* *limit F iff any open O hitting F, hits all the F_n except at most a finite number of them, and any compact K missing F, misses all the F_n except at most a finite number.* In the locally compact second countable space $\mathcal{U}$, this criterion simplifies into: *a sequence $\{F_n\}$ converges in $\mathcal{F}(\mathcal{U})$ towards a limit F iff for any point $x \in \mathcal{U}$, the sequence $\{d(x, F_n)\}$ converges in $\Re_+$ towards a limit $f(x)$ with $F = \{x : f(x) = 0\}$*, where $d(x,y)$ is any metric in $\mathcal{U}$ compatible with its topology and $d(x,F) = \mathbf{Inf}\{d(x,y),\ y \in F\}$ is the "distance" between point x and the closed set F.

From here on, we shall assume that $(\mathcal{U}, \mathcal{T})$ is locally compact and second countable, hence **metrizable.** $d(x,y)$ will denote *any* metric compatible with $\mathcal{T}$ and $d(x,F)$ (for $x \in \mathcal{U}$ and $F \in \mathcal{F}(\mathcal{U})$) will be defined as above.

The construction of the RACS theory is summarized in table 1. More details can be found in [7], [8], [9].

We recall the following fundamental theorem due to Choquet and proved by Matheron ([4]) in a probabilistic framework:

CHOQUET THEOREM. *A RACS X is entirely determined by the functional T_X defined on the set $\mathcal{K}$ of compacts of $\mathcal{U}$ by: $T_X(K) = \mathbf{Prob}(X \text{ hits } K) = \mathbf{Prob}(X \cap K \neq \emptyset), \quad \forall K \in \mathcal{K}$. Conversely, a functional T on $\mathcal{K}$ defines a (unique) RACS X verifying $\mathbf{Prob}(X \text{ hits } K) = T(K)$* **iff** *it is an alternating Choquet capacity of infinite order verifying $T(\emptyset) = 0$ and $T(K) \leq 1, \quad \forall K \in \mathcal{K}$,*

i.e. **iff** *it verifies:*

$$
\begin{array}{ll}
\textbf{(i)} & \forall K \in \mathcal{K},\ T(K) \leq 1 \\
\textbf{(ii)} & T(\emptyset) = 0 \\
\textbf{(iii)} & T(K_n) \downarrow T(K) \quad \text{whenever } K_n \downarrow K \qquad \text{(sequential continuity)} \\
\textbf{(iv)} & \forall n > 0,\ \forall (K_0, K_1, \ldots, K_n) \in \mathcal{K}^{n+1},\ S_n(K_0; K_1, \ldots, K_n) \geq 0 \quad \text{with } S_n \text{ defined by :} \\
& \left| \begin{array}{l} \qquad S_1(K_0; K_1) = T(K_0 \cup K_1) - T(K_0) \quad \text{and} \\ \forall n > 1, \quad S_n(K_0; K_1, \ldots, K_n) = S_{n-1}(K_0; K_1, \ldots, K_{n-1}) - S_{n-1}(K_0 \cup K_n; K_1, \ldots K_{n-1}) \end{array} \right.
\end{array}
\tag{1}
$$

Condition **(iv)** can be written in a more concise formula:

$$
\textbf{(iv')} \quad \forall n > 0,\ \forall (K_0, K_1, \ldots, K_n) \in \mathcal{K}^{n+1},\ T_X(K_0) \leq \sum_{\substack{I \subset \{1,\ldots,n\} \\ I \neq \emptyset}} (-1)^{|I|+1} T_X(K_0 \cup \bigcup_{i \in I} K_i)
$$

Condition (1-iii) is in fact equivalent to the *upper semi-continuity* of T on $\mathcal{K}$ ([4] p.29). T_X is called the *hitting capacity (functional)* or the *incidence capacity* of X. The term *avoidance function* was used by Kendall ([10]) to designate the functional $Q_X = 1 - T_X$.

T_X plays a role somewhat similar to cumulative probability distributions of real-valued random variables and it turns out that many properties of X can be characterized by properties of T_X (cf. table 2).

It is clear that T_X is *increasing* and attains in $\mathcal{K}$ its infimum 0 (at $\emptyset \in \mathcal{K}$), but there is no reason why it should attain its supremum $\mathbf{Sup}\,\{T(K),\ K \in \mathcal{K}\} \leq 1$. Hence, it may happen that no compact K_0 exists such that $T_X(K_0) = \mathbf{Sup}\,\{T(K),\ K \in \mathcal{K}\}$. However, when the RACS X is almost surely included in a (deterministic) compact K_0, and in particular when $\mathcal{U}$ itself is compact, T_X attains its supremum. This results from the fact that "$\mathcal{U}$ compact" implies "$\mathcal{K}(\mathcal{U}) = \mathcal{F}(\mathcal{U})$ compact" and we know that any u.s.c. function defined on a compact topological space attains its supremum.

In fact, the functional T_X can be readily extended to the set $\mathcal{B}(\mathcal{U})$ of the Borelian subsets of $\mathcal{U}$, and even to $\mathcal{P}(\mathcal{U})$ itself since T_X is u.s.c. on $\mathcal{K}$ (see [4], pp.29-30):

$$\forall B \in \mathcal{B}(\mathcal{U}), \qquad T_X^*(B) = \mathbf{Sup}\,\{T_X(K), \quad K \in \mathcal{K}, \quad K \subset B\}$$

$$\forall A \in \mathcal{P}(\mathcal{U}), \qquad T_X^*(A) = \mathbf{Inf}\,\{T_X^*(O), \quad O \in \mathcal{T}, \quad A \subset O\}$$

and T_X^* and T_X coincide on $\mathcal{K}$, so that in all cases we may write: $T_X(\mathcal{U}) = \mathbf{Sup}\,\{T_X(K),\ K \in \mathcal{K}\}$, so that T_X attains its supremum in $\mathcal{F}(\mathcal{U})$ (and not generally in $\mathcal{K} \subset \mathcal{F}(\mathcal{U})$).

Note that, even if $\mathcal{U}$ is compact, this supremum may be **strictly** smaller than 1. As a matter of fact:

$$T_X(\mathcal{U}) = \mathbf{Prob}(X \cap \mathcal{U} \neq \emptyset) = \mathbf{Prob}(X \neq \emptyset)$$

$$\text{and thus} \qquad \forall K \in \mathcal{K}\ (= \mathcal{F}), \qquad 0 \leq T_X(K) \leq \mathbf{Prob}(X \neq \emptyset) \leq 1$$

The supremum equals 1 iff X is an **a.s. non-empty** RACS ($\mathbf{Prob}(X = \emptyset) = 0$), i.e. a random variable taking its values a.s. in $\mathcal{F}'(\mathcal{U}) = \mathcal{F}(\mathcal{U}) \setminus \{\emptyset\}$. Matheron ([4]) showed that $\mathcal{F}'(\mathcal{U})$ is compact iff $\mathcal{U}$ itself is compact. If this is the case, $\{\emptyset\}$ is an isolated point in $\mathcal{F} = \mathcal{K}$.

Another functional often used in practice is the "implying functional" R_X defined on $\mathcal{K}$ by: $R_X(K) = \mathbf{Prob}(K \subset X)$. The set $\{F \in \mathcal{F},\ K \subset F\}$ is a closed set in $(\mathcal{F}(\mathcal{U}), \mathcal{T}')$, hence an event of Σ' and its probability is well defined. R_X is decreasing, with $R_X(\emptyset) = 1$.

2.4 Random Closed Sets in compact spaces

The case when $(\mathcal{U}, \mathcal{T})$ is a **compact** second countable (Hausdorff) space is of interest for the rest of this paper.

The CHOQUET theorem states that a RACS X is entirely determined by the probabilities $T_X(K) = \mathbf{Prob}(X \text{ hits } K)$ for K compact, but it is equally determined ([4], p.30) by the probabilities $T_X(O) = \mathbf{Prob}(X \text{ hits } O)$

for O open in $\mathcal{U}$, or equivalently by $Q_X(O) = \mathbf{Prob}(X \text{ misses } O)$. By noting that X misses $O \iff X \subset O^c$, we see that a RACS X is entirely determined by the probabilities $\mathbf{Prob}(X \subset F)$ for $F \in \mathcal{F}$ closed. In a compact space $\mathcal{U}$, the closed subsets are exactly the compact subsets ($\mathcal{K} = \mathcal{F}$) and the functional P_X defined by:

$$P_X(K) = \mathbf{Prob}(X \subset K), \qquad \forall K \in \mathcal{K}(= \mathcal{F}) \tag{2}$$

entirely determines X, as does T_X (or Q_X)...

The necessary and sufficient conditions for a functional P to define a RACS X (by $P_X = P$) are obviously:

$$\begin{aligned}
&\textbf{(i)} \quad \forall K \in \mathcal{K},\ P(K) \geq 0\\
&\textbf{(ii)} \quad P(\mathcal{U}) = 1\\
&\textbf{(iii)} \quad P(K_n) \downarrow P(K) \quad \text{whenever } K_n \downarrow K \qquad \text{(sequential continuity)}\\
&\textbf{(iv)} \quad \forall n > 0,\ \forall (K_0, K_1, \ldots, K_n) \in \mathcal{K}^{n+1},\ P_X(K_0) \geq \sum_{\substack{I \subset \{1,\ldots,n\} \\ I \neq \emptyset}} (-1)^{|I|+1} P_X(K_0 \cap \bigcap_{i \in I} K_i)
\end{aligned} \tag{3}$$

(P is a monotone Choquet capacity of infinite order satisfying $P(\mathcal{U}) = 1$ and $P \geq 0$). P_X is called the *inclusion capacity (functional)* of X. As for T_X, the semi-continuity of the inclusion capacity P_X allows extending it to $\mathcal{B}(\mathcal{U})$ and $\mathcal{P}(\mathcal{U})$ by setting:

$$\forall B \in \mathcal{B}(\mathcal{U}), \qquad P_X^*(B) = \mathbf{Sup}\,\{P_X(K), \quad K \in \mathcal{K}, \quad K \subset B\}$$

$$\forall A \in \mathcal{P}(\mathcal{U}), \qquad P_X^*(A) = \mathbf{Inf}\,\{P_X^*(O), \quad O \in \mathcal{T}, \quad A \subset O\}$$

so that the following equality holds:

$$\forall A \in \mathcal{P}(\mathcal{U}), \qquad T_X^*(A) + P_X^*(A^c) = 1$$

In general, we have:

$$\begin{aligned}
&\mathbf{Prob}(X \subset K) \leq \mathbf{Prob}(X \cap K \neq \emptyset) + \mathbf{Prob}(X = \emptyset)\\
&\text{ie}: \quad P_X(K) \leq T_X(K) + \mathbf{Prob}(X = \emptyset)
\end{aligned} \tag{4}$$

However, if X is a.s. non-empty, the inequality $P_X \leq T_X$ holds and it is clear that a.s. non-empty RACS play a special role here. This is emphasized by the fact that the compactness of $\mathcal{U}$ makes $\mathcal{K} \setminus \{\emptyset\}$ itself compact, and hence $\{\emptyset\}$ is *an isolated point* in $\mathcal{K} = \mathcal{F}$.

Finally, a few words about the metric nature of $\mathcal{K} \setminus \{\emptyset\}$ are in order. Let d denote any metric compatible with $\mathcal{T}$. Since $\mathcal{F}' = \mathcal{F} \setminus \{\emptyset\} = \mathcal{K}' = \mathcal{K} \setminus \{\emptyset\}$ is compact and 2nd countable, it is also metrizable. In fact, it is easy to directly construct a metric on $\mathcal{K}'$ from the distance d on $\mathcal{U}$, hence verifying directly that $\mathcal{K}'$ is metrizable:

$$\forall (K_1, K_2) \in \mathcal{K}' \times \mathcal{K}', \qquad \rho(K_1, K_2) = \mathbf{Max}\{\,\mathop{\mathbf{Sup}}_{x \in K_2} d(x, K_1), \quad \mathop{\mathbf{Sup}}_{y \in K_1} d(y, K_2)\} \tag{5}$$

ρ is called the *Hausdorff metric* and it can be verified that the topology defined by ρ on $\mathcal{K}'$ is equivalent to the relative Hit or Miss topology on $\mathcal{K}'$. This is indeed a very confortable situation and is an additional reason why we shall focus our attention on the space $\mathcal{K} \setminus \{\emptyset\}$ when $\mathcal{U}$ is compact.

Random Closed Sets in discrete finite spaces

Discrete finite spaces are compact second countable Hausdorff topological spaces; hence all the results of the previous section hold in finite spaces. The discrete topology makes **all** subsets open, closed and compact so that $\mathcal{K} = \mathcal{F} = \mathcal{O} = \mathcal{P}(\mathcal{U})$ and RACS can be simply called "Random Sets". The particularity of RACS in finite spaces is that their functionals (T_X, P_X or R_X) can be written as finite sums of the $2^{|\mathcal{U}|}$ "basic probabilities" $\mathbf{Prob}(X = A)$:

$$\forall K \subset \mathcal{U}, \quad \begin{cases} T_X(K) = \sum\limits_{\substack{A \subset \mathcal{U} \\ A \text{ hits } K}} \mathbf{Prob}(X = A) = \sum\limits_{A \cap K \neq \emptyset} \mathbf{Prob}(X = A) \\ P_X(K) = \sum\limits_{A \subset K} \mathbf{Prob}(X = A) \\ R_X(K) = \sum\limits_{K \subset A} \mathbf{Prob}(X = A) \end{cases} \tag{6}$$

2.5 Operations on RACS

Set union

It turns out that union is well suited to the RACS formalism, which comes from the fact that $\mathcal{F}(\mathcal{U})$ is stable for $\cup$ (any finite union of closed sets is closed) and that it is a *continuous* operator in $\mathcal{F}(\mathcal{U})$ equipped with the Hit or Miss topology $\mathcal{T}'$ of section 2.3. Like any other semi-continuous mappings, it is therefore measurable for the Borel σ-algebra Σ' ([7] p.82; [4] p.28), and the union $X_1 \cup X_2$ of two RACS is still a measurable mapping valued in $(\mathcal{F}(\mathcal{U}), \Sigma')$, hence a RACS of $(\mathcal{F}(\mathcal{U}), \Sigma')$. If X_1 and X_2 are two RACS of $\mathcal{U}$, we can write:

$$\forall K \in \mathcal{K}, \quad T_{X_1 \cup X_2}(K) = \mathbf{Prob}(X_1 \cup X_2 \text{ hits } K) = \mathbf{Prob}(X_1 \text{ hits } K \textbf{ or } X_2 \text{ hits } K)$$

$$= T_{X_1}(K) + T_{X_2}(K) - \mathbf{Prob}(X_1 \text{ hits } K \textbf{ and } X_2 \text{ hits } K)$$

and if X_1 and X_2 are statistically independent :

$$= T_{X_1}(K) + T_{X_2}(K) - T_{X_1}(K) \cdot T_{X_2}(K) \tag{7}$$

$$P_{X_1 \cup X_2}(K) = \mathbf{Prob}(X_1 \cup X_2 \subset K) = \mathbf{Prob}(X_1 \subset K \textbf{ and } X_2 \subset K)$$

and if X_1 and X_2 are statistically independent :

$$= P_{X_1}(K) \cdot P_{X_2}(K)$$

Since $T_{X_1 \cup X_2}(K) \geq T_{X_1}(K)$ and $T_{X_1 \cup X_2}(K) \geq T_{X_2}(K)$, we can write $T_{X_1 \cup X_2} \geq \mathbf{Max}(T_{X_1}; T_{X_2})$. Furthermore, for all $K \in \mathcal{K}$:

$$\mathbf{Prob}(X_0 \text{ hits } K \textbf{ or } X_1 \text{ hits } K) = \mathbf{Prob}(X_0 \text{ hits } K) + \mathbf{Prob}(X_1 \text{ hits } K) - \mathbf{Prob}(K_0 \text{ hits } K \textbf{ and } K_1 \text{ hits } K)$$

$$\leq \mathbf{Prob}(X_0 \text{ hits } K) + \mathbf{Prob}(X_1 \text{ hits } K)$$

so that we obtain the following bounds:

$$\begin{aligned} \mathbf{Max}(T_{X_1}; T_{X_2}) \leq T_{X_1 \cup X_2} \leq \mathbf{Min}(1; T_{X_1} + T_{X_2}) \\ \mathbf{Max}(0; P_{X_1} + P_{X_2} - 1) \leq P_{X_1 \cup X_2} \leq \mathbf{Min}(P_{X_1}; P_{X_2}) \end{aligned} \tag{8}$$

If X_1 and X_2 are independent then (8) are obviously verified. Now, consider the following extreme cases of dependence:

α) $[\forall K \in \mathcal{K}, \quad \mathbf{Prob}(X_2 \text{ hits } K \mid X_1 \text{ hits } K) = 1]$ gives $T_{X_1 \cup X_2} = T_{X_2} = \mathbf{Max}(T_{X_1}; T_{X_2})$.

β) $[\forall K \in \mathcal{K}, \quad \mathbf{Prob}(X_2 \text{ misses } K \mid X_1 \text{ hits } K) = 1]$ gives $T_{X_1 \cup X_2} = \mathbf{Min}(1; T_{X_1} + T_{X_2})$.

γ) $[\forall K \in \mathcal{K}, \quad \mathbf{Prob}(X_2 \subset K \mid X_1 \subset K) = 1]$ gives $P_{X_1 \cup X_2} = P_{X_1} = \mathbf{Min}(P_{X_1}; P_{X_2})$.

δ) $[\forall K \in \mathcal{K}, \quad \mathbf{Prob}(X_2 \subset K \mid X_1 \not\subset K) = 1]$ gives $P_{X_1 \cup X_2} = P_{X_1} + P_{X_2} - 1 = \mathbf{Max}(0; P_{X_1} + P_{X_2} - 1)$.

If X_1 and X_2 are two RACS with respective hitting functionals T_{X_1} and T_{X_2} and respective inclusion functionals P_{X_1} and P_{X_2}, neither $\mathbf{Max}(T_{X_1}; T_{X_2})$, $\mathbf{Min}(1; T_{X_1} + T_{X_2})$, $\mathbf{Min}(P_{X_1}; P_{X_2})$ nor $\mathbf{Max}(0; P_{X_1} + P_{X_2} - 1)$ define the hitting or inclusion functional of a RACS in general (these functionals are not Choquet capacities) *unless* the statistical dependence between X_1 and X_2 is of one of the 4 types above.

However, even if the statistical dependence between X_1 and X_2 does not correspond to any of the extreme cases above, it is still possible for the capacity functionals of the combined RACS $X_1 \cup X_2$ to attain one bound of (8) *on the set $\mathcal{S}(\mathcal{U})$ of the singletons of $\mathcal{U}$* ($\mathcal{S}(\mathcal{U}) \subset \mathcal{K}(\mathcal{U})$ since $\mathcal{U}$ is Hausdorff). In fact, one can easily construct two RACS X_1 and X_2 such that $T_{X_1 \cup X_2}$ (resp. $P_{X_1 \cup X_2}$) is constrained on $\mathcal{S}(\mathcal{U})$ to verify any relation within the two boundaries of (8). This is clear from condition (**iv'**) of the Choquet theorem: for any given set of values $\{T_{X_1 \cup X_2}(\{k\})\}_{k \in \mathcal{U}}$ (resp. $\{P_{X_1 \cup X_2}(\{k\})\}_{k \in \mathcal{U}}$) verifying (8), it is always possible to choose the values of $T_{X_1 \cup X_2}$ (resp. $P_{X_1 \cup X_2}$) on the non-singleton compact subsets of $\mathcal{U}$ (including all non-singleton finite subsets) so that the Choquet theorem is verified.

Finally, note that the union of two a.s. non-empty RACS is still a.s. non-empty.

<u>Set intersection</u>

Like union, intersection has good properties that make it well suited to the RACS theory: $\mathcal{F}(\mathcal{U})$ is stable for $\cap$ and in $\mathcal{F}(\mathcal{U})$ equipped with the Hit or Miss topology $\mathcal{T}'$, $\cap$ is *upper semi-continuous* (but not continuous, see [7] p.77, [4] p.7), hence measurable for the Borel σ-algebra Σ', so that the intersection $X_1 \cap X_2$ of two RACS is still a RACS of $(\mathcal{F}(\mathcal{U}), \Sigma')$. Unfortunately, there is no simple way of writing the hitting and including functionals of $X_1 \cap X_2$ in terms of those of X_1 and X_2, as is the case with union, even if X_1 and X_2 are assumed statistically independent. In general, we only have the following inequalities:

$$
\begin{aligned}
0 \leq T_{X_1 \cap X_2} \leq \mathbf{Min}(T_{X_1}; T_{X_2}) \\
\mathbf{Max}(P_{X_1}; P_{X_2}) \leq P_{X_1 \cap X_2} \leq 1
\end{aligned}
\tag{9}
$$

With the implying functional $R_{X_1 \cap X_2}$ however, we get the following stronger inequalities:

$$
\mathbf{Max}(0; R_{X_1} + R_{X_2} - 1) \leq R_{X_1 \cap X_2} \leq \mathbf{Min}(R_{X_1}; R_{X_2}) \tag{10}
$$

And when $K = \{k\}$ is a singleton, $T_X(\{k\}) = R_X(\{k\})$ and we can write:

$$
\mathbf{Max}(0; T_{X_1}(\{k\}) + T_{X_2}(\{k\}) - 1) \leq T_{X_1 \cap X_2}(\{k\}) \leq \mathbf{Min}(T_{X_1}(\{k\}); T_{X_2}(\{k\})) \tag{11}
$$

which is obviously verified when X_1 and X_2 are statistically independent. Furthermore, one can always construct two RACS X_1 and X_2 such that $T_{X_1 \cap X_2}$ is constrained on $\mathcal{S}(\mathcal{U})$ to verify any relation within the two boundaries of (10) (including the boundaries themselves).

In a *finite* space $\mathcal{U}$ (equipped with the discrete topology), we can write:

$$
\forall K \in \mathcal{K} \setminus \{\emptyset\}, \qquad
\begin{aligned}
T_{X_1 \cap X_2}(K) &= \sum_{A \text{ hits } K} \Big(\sum_{B \cap C = A} \mathbf{Prob}(X_1 = B;\ X_2 = C) \Big) \\
P_{X_1 \cap X_2}(K) &= \sum_{A \subset K} \Big(\sum_{B \cap C = A} \mathbf{Prob}(X_1 = B;\ X_2 = C) \Big) \\
R_{X_1 \cap X_2}(K) &= \sum_{K \subset A} \Big(\sum_{B \cap C = A} \mathbf{Prob}(X_1 = B;\ X_2 = C) \Big)
\end{aligned}
\tag{12}
$$

Of course, there is no reason why $X_1 \cap X_2$ should be a.s. non-empty (or even non a.s. empty) even if X_1 and X_2 are both a.s. non-empty.

Probabilistic combinations

A common property of all combinations based on set-theoretic operators is that they are well suited to the management of *imprecision* since imprecision is a set-theoretic concept.

However, by construction, *these combinations cannot deal with problems of uncertainty.* This is clear when we notice that they are *fatal* in the sense that any piece of evidence will irremediably alter the Representation of the World: no matter what evidence and how much information is subsequently provided, the Representation will never go back to its previous state. If the all the evidence is somewhat *uncertain*, this over-optimistic way of accumulating it will generally fail. If we want to deal with uncertain data, we must select a *probabilistic combination* operator, since uncertainty is a probabilistic concept.

The simplest probabilistic operation one can think of is the *mean* operation. If $X_1, \ldots, X_n$ are n statistically independent RACS and $(\alpha_1, \ldots, \alpha_n) \in (\Re^+)^n \setminus \{(0, \ldots, 0)\}$ are n real positive numbers at least one of which is non-zero, the functional T defined on $\mathcal{K}$ by:

$$\forall K \in \mathcal{K}, \quad T(K) = \frac{\sum_{i=1}^{n} \alpha_i \cdot T_{X_i}(K)}{\sum_{i=1}^{n} \alpha_i} \tag{13}$$

obviously verifies the conditions of the Choquet theorem, and hence defines a RACS of $\mathcal{U}$, called *mean RACS* of $X_1, \ldots, X_n$ with *weights* $\alpha_1, \ldots, \alpha_n$. The corresponding inclusion capacity functional is:

$$\forall K \in \mathcal{K}, \quad P(K) = \frac{\sum_{i=1}^{n} \alpha_i \cdot P_{X_i}(K)}{\sum_{i=1}^{n} \alpha_i} \tag{13'}$$

This is truly a probabilistic accumulation since we do not attempt to combine the realization sets themselves (using set-theoretic operators), but merely take the union of the realization spaces of the random variable. A direct consequence is that *combination is not fatal* in the sense that the relative weight of any piece of evidence will decrease as more and more information is gathered, and eventually become negligible if the amount of received information is large enough.

3. Mathematical connections with other theories

3.1 upper and lower probabilities

3.1.1 induced by a multi-valued mapping

This approach dates back to [11] and originated from the idea that a *multivalued mapping from a space X to a space S carries a probability measure defined over the subsets of X into a system of upper and lower probabilities over subsets of S* ([11]). See [5] for a more recent formulation of the same idea.

Let $(\Omega, \Sigma_\Omega, \mathbf{Prob})$ be a probability space and $(\mathcal{U}, \Sigma_\mathcal{U})$ a measurable space. Let $\Gamma : \Omega \longrightarrow \mathcal{P}(\mathcal{U})$ be a "multivalued mapping" over $\mathcal{U}$, i.e. a mapping taking values in $\mathcal{P}(\mathcal{U})$. Consider the two following "inverses" of Γ:

$$\begin{aligned} &\text{lower inverse} \quad \Gamma_* : \mathcal{P}(\mathcal{U}) \longrightarrow \mathcal{P}(\Omega) \\ &\qquad\qquad A \longmapsto \Gamma_*(A) = \{\omega \in \Omega;\ \Gamma(\omega) \subset A\} \\ &\text{upper inverse} \quad \Gamma^* : \mathcal{P}(\mathcal{U}) \longrightarrow \mathcal{P}(\Omega) \\ &\qquad\qquad A \longmapsto \Gamma^*(A) = \{\omega \in \Omega;\ \Gamma(\omega) \text{ hits } A\} \end{aligned} \tag{14}$$

(obviously $\Gamma_*(\mathcal{U}) = \Omega$, $\Gamma^*(\emptyset) = \emptyset$ and $\Gamma_*(A^c) = (\Gamma^*(A))^c$). We say that Γ is *strongly measurable* with respect to Σ_Ω and $\Sigma_\mathcal{U}$ whenever the lower and upper-inverses of any event of $\Sigma_\mathcal{U}$ are events of Σ_Ω. Given a strongly measurable (multivalued) mapping Γ, define the *lower-probability* P_* and the *upper-probability* P^* induced by Γ, by:

$$\begin{aligned} \forall A \in \Sigma_\mathcal{U}, \quad P_*(A) &= \mathbf{Prob}(\Gamma_*(A)) \quad \text{noted} \quad \mathbf{Prob}(\Gamma \subset A) \\ P^*(A) &= \mathbf{Prob}(\Gamma^*(A)) \quad \text{noted} \quad \mathbf{Prob}(\Gamma \text{ hits } A) \end{aligned} \tag{15}$$

Note that neither P_* nor P^* are *measures* in general (they are not necessarily additive) but they are *capacities.* We obviously have: $P_*(\mathcal{U}) = 1$, $P^*(\emptyset) = 0$ and $P_*(A^c) = 1 - P^*(A), \forall A \in \Sigma_\mathcal{U}$.

The relationship with the RACS theory is clear: in the latter, we explicitly construct a σ-algebra on $\mathcal{F}(\mathcal{U}) \subset \mathcal{P}(\mathcal{U})$ by topologizing this space with the Hit or Miss topology induced by a topology on $\mathcal{U}$. Then we can legitimately speak of the *measurability* of a mapping $\Gamma : \Omega \longrightarrow \mathcal{F}(\mathcal{U})$ without recourse to the somewhat artificial concept of "strong measurability". Although the above approach is more general (no topology is assumed on $\mathcal{U}$), we argued (2.2) that there is no practical use for such abstract theories and that assuming a topological structure on $\mathcal{U}$ and restricting our study to the mappings valued in $\mathcal{F}(\mathcal{U})$ are the minimum concessions that physicists are entitled to ask from mathematicians... As soon as a topology on $\mathcal{U}$ is assumed ($\Sigma_\mathcal{U}$ being its Borel σ-algebra) and Γ takes its values in $\mathcal{F}(\mathcal{U})$, both approaches are strictly equivalent, i.e. Γ is strongly measurable w.r.t. Σ_Ω and $\Sigma_\mathcal{U}$ iff it is measurable w.r.t. Σ_Ω and $\Sigma' = \Sigma_{\mathcal{F}(\mathcal{U})}$.

Given a system of lower and upper probabilities P_* and P^* induced by a strongly measurable (multivalued) mapping Γ such that the event $\Gamma_*(\emptyset) = \Gamma^*(\mathcal{U})$ has zero probability measure, consider the set $\mathcal{C}$ of all (genuine, i.e. σ-additive) probability measures P that can be defined on the measurable space $(\mathcal{U}, \Sigma_\mathcal{U})$ and which verify:

$$\forall A \in \Sigma_\mathcal{U}, \quad P_*(A) \leq P(A) \leq P^*(A)$$

$\mathcal{C}$ is clearly a *convex closed set of probability measures* ([5], [12]). For a given event $A \in \Sigma_\mathcal{U}$, consider the probability measures $\check{P}$ and $\hat{P}$ defined respectively by $\check{P}(A) = P_*(A)$, $\check{P}(A^c) = 1 - P_*(A)$ and $\hat{P}(A) = P^*(A)$, $\hat{P}(A^c) =$

$1 - P^*(A)$. We have:

$$P_*(A) = \check{P}(A) \leq P^*(A)$$

$$P_*(A^c) \leq \check{P}(A^c) = P^*(A^c)$$

$$P_*(A) \leq \hat{P}(A) = P^*(A)$$

$$P_*(A^c) = \hat{P}(A^c) \leq P^*(A^c)$$

and therefore $\check{P} \in \mathcal{C}$ and $\hat{P} \in \mathcal{C}$, and obviously $\check{P}(A) = \underset{P \in \mathcal{C}}{\mathrm{Inf}}(P(A))$ and $\hat{P}(A) = \underset{P \in \mathcal{C}}{\mathrm{Sup}}(P(A))$ so that we can conclude:

$$\forall A \in \Sigma_{\mathcal{U}}, \qquad P_*(A) = \underset{P \in \mathcal{C}}{\mathrm{Inf}}\, P(A)$$

$$P^*(A) = \underset{P \in \mathcal{C}}{\mathrm{Sup}}\, P(A)$$

3.1.2 induced by a convex subset of probability measures

Conversely, consider a class $\mathcal{C}$ of (regular σ-additive) probability measures over a measurable space $(\mathcal{U}, \Sigma_{\mathcal{U}})$ and define:

$$\begin{aligned} &\text{lower probability}: \quad \forall A \in \Sigma_{\mathcal{U}}, \quad P_*(A) = \underset{P \in \mathcal{C}}{\mathrm{Inf}}\, P(A) \\ &\text{upper probability}: \quad \forall A \in \Sigma_{\mathcal{U}}, \quad P^*(A) = \underset{P \in \mathcal{C}}{\mathrm{Sup}}\, P(A) \end{aligned} \tag{16}$$

(we obviously have $P_*(\mathcal{U}) = P^*(\mathcal{U}) = 1$, $P_*(\emptyset) = P^*(\emptyset) = 0$ and $P_*(A^c) = 1 - P^*(A)$, $\forall A \in \Sigma_{\mathcal{U}}$). Since the same lower and upper probabilities are yielded by the convex closure of $\mathcal{C}$ as by $\mathcal{C}$ itself, we may as well restrict $\mathcal{C}$ to be a convex closed subset of measures. These sets are called *belief structures* in [13] and are thought of as the sets of the possible bets of a rational gambler.

The main point here is that there does not necessarily exist a multivalued mapping Γ that "fits" a given class $\mathcal{C}$ according to equations (15). In order to guaranty this existence and hence the equivalence of the two definitions (15) and (16), class $\mathcal{C}$ must verify the additional following constraint (a proof in general compact metrizable spaces is given in the appendix, Proposition 2):

$$\forall n \geq 1, \quad \forall (A_1, \ldots, A_n) \in (\Sigma_{\mathcal{U}})^n, \quad \forall P \in \mathcal{C}, \qquad P(\bigcup_{i=1}^{n} A_i) \geq \sum_{I \subset \{1,\ldots,n\}} (-1)^{|I|+1} \underset{P \in \mathcal{C}}{\mathrm{Inf}}(P(\bigcap_{i \in I} A_i)) \tag{17}$$

or equivalently:

$$\forall n \geq 1, \quad \forall (A_1, \ldots, A_n) \in (\Sigma_{\mathcal{U}})^n, \quad \forall P \in \mathcal{C}, \qquad P(\bigcap_{i=1}^{n} A_i) \leq \sum_{I \subset \{1,\ldots,n\}} (-1)^{|I|+1} \underset{P \in \mathcal{C}}{\mathrm{Sup}}(P(\bigcup_{i \in I} A_i)) \tag{17'}$$

Now, given *any* convex closed set $\mathcal{C}$ of probability measures on $\Sigma_{\mathcal{U}}$, it is of interest to determine a RACS X such that:

$$\forall A \in \Sigma_{\mathcal{U}}, \forall P \in \mathcal{C}, \quad P_X(A) \leq P(A) \leq T_X(A)$$

The "smallest" such RACS (i.e. the RACS X with the largest P_X and the smallest T_X) is obviously given by the *closure C' of C w.r.t. inequalities (17)*, i.e. the smallest set C' that includes the convex closed set C and that verifies inequalities (17). Indeed, the closure w.r.t. (17) of a closed convex subset of probability measures remains a closed convex subset of probability measures. Then we have, by construction (P_* and P^* being defined as in (16)):

$$0 \leq P_X \leq P_* \leq P^* \leq T_X \leq 1 \tag{18}$$

but the inequality $0 \leq P_*(A) \leq P(A) \leq P^*(A) \leq 1$ does not hold in C' (for all $A \in \Sigma_{\mathcal{U}}$) in general.

3.2 Dempster-Shafer (DS) Theory of Evidence

Let $\mathcal{U}$ denote the *finite* Universe (called Frame of Discernment) and $2^{\mathcal{U}} = \mathcal{P}(\mathcal{U})$ the power set of $\mathcal{U}$ which is also finite ($|2^{\mathcal{U}}| = 2^{|\mathcal{U}|}$).

3.2.1 Belief/Plausibility functions

A **Belief function** is a function from $2^{\mathcal{U}}$ into the unit real interval $[0;1]$ that verifies:

$$\left|\begin{array}{ll} \text{(i)} & Bel(\emptyset) = 0 \\ \text{(ii)} & Bel(\mathcal{U}) = 1 \\ \text{(iii)} & Bel(\bigcup_{i=1,\dots,n} A_i) \geq \sum_{\substack{I \subset \{1,\dots,n\} \\ I \neq \emptyset}} (-1)^{|I|+1} Bel(\bigcap_{i \in I} A_i) \quad \text{for every finite family } A_1, \dots, A_n \end{array}\right. \tag{19}$$

A direct consequence of these conditions is: $\forall A \subset \mathcal{U}, \quad Bel(A) + Bel(A^c) \leq 1$. **Plausibility functions**, noted Pls, are defined by: $Pls(A) = 1 - Bel(A^c), \quad \forall A \subset \mathcal{U}$. Belief and Plausibility functions play a dual role in the theory and it is clear that a function $Pls : 2^{\mathcal{U}} \longrightarrow [0;1]$ is a Plausibility function iff it verifies:

$$\left|\begin{array}{ll} \text{(i)} & Pls(\emptyset) = 0 \\ \text{(ii)} & Pls(\mathcal{U}) = 1 \\ \text{(iii)} & Pls(\bigcap_{i=1,\dots,n} A_i) \leq \sum_{\substack{I \subset \{1,\dots,n\} \\ I \neq \emptyset}} (-1)^{|I|+1} Pls(\bigcup_{i \in I} A_i) \quad \text{for every finite family } A_1, \dots, A_n \end{array}\right. \tag{20}$$

Note that conditions (19-iii) and (20-iii) are both loosened versions of the classical *Poincaré formulas* that hold for any measure μ on Σ:

$$\mu(\bigcup_{i=1,\dots,n} A_i) = \sum_{\substack{I \subset \{1,\dots,n\} \\ I \neq \emptyset}} (-1)^{|I|+1} \mu(\bigcap_{i \in I} A_i) \qquad \forall n > 0 \text{ and } \forall A_1, \dots, A_n \in \Sigma \tag{21}$$

$$\mu(\bigcap_{i=1,\ldots,n} A_i) = \sum_{\substack{I\subset\{1,\ldots,n\}\\ I\neq\emptyset}} (-1)^{|I|+1}\mu(\bigcup_{i\in I} A_i) \qquad \forall n>0 \text{ and } \forall A_1,\ldots,A_n \in \Sigma \tag{21'}$$

The **Belief interval** associated with a subset A of $\mathcal{U}$ is simply:

$$[Bel(A); Pls(A)] = [Bel(A); 1 - Bel(A^c)] = [1 - Pls(A^c); Pls(A)] \subset [0;1]$$

The Belief interval associated with A summarizes the information about A and its complement in $\mathcal{U}$; its length $Pls(A) - Bel(A)$ represents the amount of belief that is neither commited to A nor to its complement, i.e. the amount of *ignorance* about A and its complement. A **Mass function** is a function from $2^{\mathcal{U}}$ into $[0,1]$ that verifies:

$$\left|\begin{array}{ll} \text{(i)} & m(\emptyset) = 0 \\ \text{(ii)} & \sum_{A\subset\mathcal{U}} m(A) = 1 \end{array}\right. \tag{22}$$

Given a Mass function m over $\mathcal{U}$, a **Communality function** can be defined by setting $q(A) = \sum_{A\subset B} m(B)$.

Mass functions are related to Belief, Plausibility and Communality functions by the following equations:

$$\forall A \in 2^{\mathcal{U}}, \quad \begin{array}{ll} Bel(A) = \sum_{B\subset A} m(B) & m(A) = \sum_{B\subset A} (-1)^{|A-B|} Bel(B) \\ Pls(A) = 1 - \sum_{B\subset A^c} m(B) = \sum_{B\cap A\neq\emptyset} m(B) & m(A) = \sum_{B\supset A^c} (-1)^{|A\cap B|+1} Pls(B) \\ q(A) = \sum_{A\subset B} m(B) & m(A) = \sum_{A\subset B} (-1)^{|B-A|} q(B) \end{array} \tag{23}$$

Hence, the 4 fundamental concepts of the theory (Mass, Belief, Plausibility and Communality functions) are all "equivalent" in the sense that any of them is sufficient to define the other three.

The subsets $A \subset \mathcal{U}$ such that $m(A) > 0$ are called the *focal elements* of m. The focal elements of a Belief function *Bel* (or a Plausibility function *Pls*) are simply the focal elements of the associated Mass function m. We say that a Belief function (or equivalently a Plausibility function) is *consonant* whenever its focal elements are *nested*, i.e. they can be totally ordered (by $\subset$). A **Bayesian Belief function** is a Belief function whose only focal elements are *singletons* of $\mathcal{U}$. Refer to [14] for an application of the DS theory to a concrete AI problem.

In order to compare the DS representation with Belief/Plausibility functions and the RACS theory, we have no alternative but to consider spaces where both are well defined, i.e. where the axioms of both are satisfied: **discrete finite spaces** are the only such sets, as they are the only finite topological spaces that are Hausdorff, compact and second countable (figure 1).

In the appendix (Proposition 1), we prove the following equivalence:

Let f be a function defined on the power set $\mathcal{P}(\mathcal{U})$ of a set $\mathcal{U}$.

Then f verifies:

$$\forall n > 0,\ \forall (A_0, A_1, \ldots, A_n) \in \mathcal{P}(\mathcal{U})^{n+1},\ f(A_0) \leq \sum_{\substack{I \subset \{1,\ldots,n\} \\ I \neq \emptyset}} (-1)^{|I|+1} f(A_0 \cup \bigcup_{i \in I} A_i) \tag{24}$$

iff *f is* **increasing** *and verifies:*

$$\forall n > 0,\ \forall (A'_1, \ldots, A'_n) \in \mathcal{P}(\mathcal{U})^n,\ f(\bigcap_{i=1,\ldots,n} A'_i) \leq \sum_{\substack{I \subset \{1,\ldots,n\} \\ I \neq \emptyset}} (-1)^{|I|+1} f(\bigcup_{i \in I} A'_i) \tag{25}$$

Let $\mathcal{U}$ be a *finite* set, equipped with the discrete topology. Using the above equivalence, it is not difficult to show that:

The Plausibility functions Pls on $\mathcal{U}$ are exactly the alternating Choquet capacities of infinite order satisfying $Pls(\emptyset) = 0$ and $Pls(\mathcal{U}) = 1$.

(the semi-continuity condition is trivially verified since every function defined on a discrete space is continuous)

And since a functional T defined on $\mathcal{K}(\mathcal{U}) = \mathcal{P}(\mathcal{U})$ entirely determines a RACS on $\mathcal{U}$ iff it is an alternating capacity of infinite order verifying $T(\emptyset) = 0$ and $T \leq 1$ (CHOQUET theorem), we conclude:

The Plausibility functions Pls defined on a (finite) Frame of Discernment $\mathcal{U}$ are exactly the hitting capacities T_X of the almost surely non-empty Random (Closed) Sets X of the discrete topological space $\mathcal{U}$.

By changing f into $1 - f^c$, we get the dual proposition:

The Belief functions Bel on $\mathcal{U}$ are exactly the monotone Choquet capacities of infinite order satisfying $Bel(\emptyset) = 0$ and $Bel(\mathcal{U}) = 1$.

and consequently:

The Belief functions Bel defined on a (finite) Frame of Discernment $\mathcal{U}$ are exactly the inclusion capacities P_X of the almost surely non-empty Random (Closed) Sets X of the discrete topological space $\mathcal{U}$.

The Communality functions q are obviously the implying functionals R_X of RACS X, as can be seen by comparing equations (6) and (23). As for Mass functions m, they are equivalent to RACS *probability densities* $f = \frac{dP'}{d\nu}$ where ν is the counting measure in the (finite) space $\mathcal{P}(\mathcal{U})$:

$$\forall A \in \mathcal{P}(\mathcal{U}), \quad m(A) = \mathbf{Prob}(X = A) \tag{26}$$

In light of this equivalence, the link between the DS and the upper/lower probability formalisms (3.1) is clear: a RACS of $\mathcal{U}$ being a $(\Sigma_\Omega, \Sigma_\mathcal{U})$-strongly measurable multivalued mapping, definitions (14) and (15) are valid and we have: $P_X = P_* = Bel$ and $T_X = P^* = Pls$. These functionals can be equally viewed as the lower and upper probabilities induced by the convex closed set of measures $\mathcal{C} = \{P$ probability measure on $\Sigma_\mathcal{U}$; $\forall A \in \Sigma_\mathcal{U},\ Bel(A) \leq P(A) \leq Pls(A)\}$ since $\mathcal{C}$ obviously verifies (17).

3.2.2 Dempster's rule

Hence, the Dempster-Shafer formalism deals exclusively with a.s. non-empty RACS. The problem with RACS intersection is that $\mathcal{F}(\mathcal{U}) \setminus \{\emptyset\}$ is not stable for $\cap$, so that the intersection of two a.s. non-empty RACS need not be a.s. non-empty. This is most unfortunate as empty sets do not "carry" any information (at least in the original DS framework). Thus, we must find a way to obtain an a.s. non-empty RACS by combining two a.s. non-empty RACS X_1 and X_2 using intersection. The answer is simple: if the event $X_1 \cap X_2 \neq \emptyset$ (this is indeed an event of Σ') has a non-zero probability, we can consider the *conditional RACS $X_1 \cap X_2$ given this event*, and this will be a.s. non-empty by construction. However,if $\mathbf{Prob}(X_1 \cap X_2 = \emptyset) = 1$, then it is not possible to combine the two RACS using $\cap$ in such a way that the result is an a.s. non-empty RACS. In this latter case, we simply state that *"the two pieces of evidence represented by RACS X_1 and X_2 are flatly (or totally) conflicting"* and do not attempt to combine them. Suppose that this is not the case however. A simple calculation gives, for any non-empty (compact) subset K:

$$T_{X_1 \oplus X_2}(K) = \mathbf{Prob}(X_1 \cap X_2 \text{ hits } K \mid X_1 \cap X_2 \neq \emptyset) = \frac{T_{X_1 \cap X_2}(K)}{\mathbf{Prob}(X_1 \cap X_2 \neq \emptyset)} = T_{X_1 \cap X_2 | X_1 \cap X_2 \neq \emptyset}(K) \tag{27}$$

The amount $\kappa = \mathbf{Prob}(X_1 \cap X_2 = \emptyset)$ is called the *amount of conflict* between the two RACS X_1 and X_2: it is simply the probability that the two RACS may be disjoint and is thought of as a measure of the "conflict" between the two pieces of evidence represented by X_1 and X_2.

In the finite case, we can write, for all $K \in \mathcal{K} \setminus \{\emptyset\}$:

$$\begin{aligned} T_{X_1 \oplus X_2}(K) = \frac{\sum_{A \text{ hits } K} \mathbf{Prob}(X_1 \cap X_2 = A)}{\mathbf{Prob}(X_1 \cap X_2 \neq \emptyset)} &= \frac{\sum_{A \text{ hits } K} \left(\sum_{B \cap C = A} \mathbf{Prob}(X_1 = B;\ X_2 = C) \right)}{\sum_{A \cap B \neq \emptyset} \mathbf{Prob}(X_1 = A;\ X_2 = B)} \\ &= \frac{\sum_{A \text{ hits } K} \left(\sum_{B \cap C = A} \mathbf{Prob}(X_1 = B;\ X_2 = C) \right)}{1 - \sum_{A \cap B = \emptyset} \mathbf{Prob}(X_1 = A;\ X_2 = B)} \end{aligned} \tag{28}$$

In the particular case when X_1 and X_2 are assumed *statistically independent*, we may write:

$$\forall K \in \mathcal{K} \setminus \{\emptyset\}, \qquad T_{X_1 \oplus X_2}(K) = \frac{\sum_{A \text{ hits } K} \left(\sum_{B \cap C = A} \mathbf{Prob}(X_1 = B) \cdot \mathbf{Prob}(X_2 = C) \right)}{1 - \sum_{A \cap B = \emptyset} \mathbf{Prob}(X_1 = A) \cdot \mathbf{Prob}(X_2 = B)} \tag{29}$$

This operator $\oplus$ is known as **Dempster's rule of combination** ([1]). Remembering the equivalence relations between the RACS and the DS formalisms (3.2.1), we can write:

$$\forall K \in \mathcal{K}\setminus\{\emptyset\}, \qquad Pls_1 \oplus Pls_2(K) = \frac{\sum\limits_{A \text{ hits } K}\left(\sum\limits_{B\cap C=A} m_1(B)\cdot m_2(C)\right)}{1-\sum\limits_{A\cap B=\emptyset} m_1(A)\cdot m_2(B)} \tag{30}$$

where m_1 (resp. m_2) is the Mass function of Pls_1 (resp. Pls_2). It is clear that Dempster's rule is both *commutative* and *associative*, allowing us to define such quantities as: $m_1 \oplus m_2 \oplus \cdots \oplus m_n$, which is independent of the order of combination:

$$\bigoplus_{i=1,\dots,n} m_i \quad \begin{cases} [\bigoplus\limits_{i=1,\dots,n} m_i](\emptyset) = 0 \\ [\bigoplus\limits_{i=1,\dots,n} m_i](A) = \dfrac{\sum\limits_{A=A_1\cap\cdots\cap A_n} m_1(A_1)\cdots m_n(A_n)}{1-\sum\limits_{A_1\cap\cdots\cap A_n=\emptyset} m_1(A_1)\cdots m_n(A_n)} & \forall A \neq \emptyset \end{cases}$$

Implying functionals (or their DS counterpart, Communality functions) allow a very concise definition of Dempster's rule:

$$R_{X_1\oplus X_2}(K) = \mathbf{Prob}(K \subset X_1\cap X_2 \mid X_1\cap X_2 \neq \emptyset) = \frac{R_{X_1\cap X_2}(K)}{\mathbf{Prob}(X_1\cap X_2\neq\emptyset)} = \frac{R_{X_1}(K)\cdot R_{X_2}(K)}{\mathbf{Prob}(X_1\cap X_2\neq\emptyset)}$$

the rightmost equality being verified if X_1 and X_2 are independent.

3.3 Possibility theory

3.3.1 Zadeh's Possibility/Necessity measures

Zadeh ([15]) defined a *possibility measure* on $\mathcal{U}$ as a set-functional Π valued in $[0;1]$ and verifying:

$$\begin{array}{|ll} \text{(i)} & \Pi(\emptyset) = 0 \\ \text{(ii)} & \Pi(\mathcal{U}) = 1 \\ \text{(iii)} & \forall (A,B)\in\mathcal{P}(\mathcal{U})^2, \quad \Pi(A\cup B) = \mathbf{Max}(\Pi(A),\Pi(B)) \end{array} \tag{31}$$

A possibility measure Π has the property that it is entirely determined by its *possibility distribution* $\phi : \mathcal{U} \longrightarrow [0;1]$ defined by $\phi(x) = \Pi(\{x\})$. Indeed, we have: $\forall A \subset \mathcal{U}, \quad \Pi(A) = \mathbf{Sup}\{\phi(x);\ x\in A\}$. Note however that there may exist other functions $\phi' : \mathcal{U} \longrightarrow [0;1]$ such that $\forall A \subset \mathcal{U}, \quad \Pi(A) = \mathbf{Sup}\{\phi'(x);\ x\in A\}$.

Conversely, starting from a function $\phi : \mathcal{U} \longrightarrow [0;1]$, it defines a possibility measure by $\forall A \subset \mathcal{U}, \quad \Pi(A) = \mathbf{Sup}\{\phi(x);\ x\in A\}$ iff ϕ is *normalized*, i.e. there exists some $x_0 \in \mathcal{U}$ such that $\phi(x_0) = 1$.

If Π is a possibility measure, it is customary to call the functional $\aleph : \ \aleph(A) = 1 - \Pi(A^c)$ a *necessity measure.*

Necessity measures can be axiomatically defined by:

$$\left| \begin{array}{l} \text{(i)} \quad \aleph(\emptyset) = 0 \\ \\ \text{(ii)} \quad \aleph(\mathcal{U}) = 1 \\ \\ \text{(iii)} \quad \forall (A,B) \in \mathcal{P}(\mathcal{U})^2, \quad \aleph(A \cap B) = \mathbf{Min}(\aleph(A), \aleph(B)) \end{array} \right. \qquad (32)$$

If $\mathcal{U}$ is finite, it is clear that *consonant* Plausibility (resp. Belief) functions are possibility (resp. necessity) measures. It can be shown directly (appendix, Proposition 3) that *every* possibility (resp. necessity) measure, as defined by (31) (resp. (32)), verifies (20) (resp. (19)) and hence is a Plausibility (resp. Belief) function (in spite of what [16] suggests).

To extend the argument to infinite spaces, we must assume a topological structure on $\mathcal{U}$ and for the reasons given in 2.3, we assume that $\mathcal{U}$ is locally compact 2nd countable Hausdorff. This topological setting would be quite useless if we do not impose any (semi) continuity requirement on the possibility (and necessity) measures that can be defined on the Borel σ-algebra $\Sigma_{\mathcal{U}}$. This is done by strengthening conditions (31-iii) and (32-iii) into:

$$(31 - \text{iii}') \qquad \Pi(\bigcup_{i \in I} A_i) = \mathop{\mathbf{Sup}}_{i \in I} \Pi(A_i) \quad \text{for any countable family of events } A_i \in \Sigma_{\mathcal{U}}$$

$$(32 - \text{iii}') \qquad \aleph(\bigcap_{i \in I} A_i) = \mathop{\mathbf{Inf}}_{i \in I} \aleph(A_i) \quad \text{for any countable family of events } A_i \in \Sigma_{\mathcal{U}}$$

Indeed, this implies that $\Pi(O_n) \uparrow \Pi(O)$ whenever $O_n \uparrow O$, i.e. Π is l.s.c. on $\mathcal{O}$ (since it is increasing), or equivalently u.s.c. on $\mathcal{K}$. Similarly, $\aleph$ is such that $\aleph(K_n) \downarrow \aleph(K)$ whenever $K_n \downarrow K$ and hence is u.s.c. on $\mathcal{K}$ (since it is increasing). Furthermore, among the functions ϕ such that $\forall A \subset \mathcal{U}, \quad \Pi(A) = \mathbf{Sup}\{\phi(x);\ x \in A\}$ and $\aleph(A) = \mathbf{Inf}\{1 - \phi(x);\ x \in A^c\}$, the one defined by $\phi(x) = \Pi(\{x\})$ is particularized as being the only one which is u.s.c. ($1 - \phi$ being l.s.c.).

If $\mathcal{U}$ is compact, it is clear that every possibility measure Π (resp. necessity measure $\aleph$) defines an a.s. non-empty RACS X by $T_X = \Pi$ on $\mathcal{K}$ (resp. $P_X = \aleph$ on $\mathcal{K}$). This can be verified directly with the Choquet theorem, but it is also a consequence of a more general relationship between the RACS theory and Fuzzy sets (Π and $\aleph$ define the canonical RACS $\kappa(S)$ associated with the fuzzy set S represented by the membership function ϕ, and since ϕ is normalized, $\kappa(S)$ is a.s. non-empty; see 3.4.1).

3.3.2 Giles' generalized theory

Giles ([13]) proposed a generalization of the concept of Possibility/Necessity, interpreted in the context of betting behaviour by a rational gambler. His theory can be axiomatically defined by the single property (adapted from [13]):

$\forall n \geq 1, \quad \forall r \in \{0, \dots, n\}, \quad \forall (A_1, \dots, A_n) \in \mathcal{P}(\mathcal{U})^n,$

$$\bigcup_{\substack{I \subset \{1,\dots,n\} \\ |I| = r}} (\bigcap_{i \in I} A_i) = \mathcal{U} \quad \Longrightarrow \quad \forall s \in \{0, \dots, n-r\}, \quad \sum_{i=1,\dots,n} \Pi(A_i) \geq r + s \cdot \Pi(\bigcup_{\substack{I \subset \{1,\dots,n\} \\ |I| = r+s}} (\bigcap_{i \in I} A_i)) \qquad (33)$$

which, as a consequence, yields a relaxed version of (31-iii):

$$\forall A, B, \quad A \cap B = \emptyset \Longrightarrow \mathbf{Max}(\Pi(A), \Pi(B)) \leq \Pi(A \cup B) \leq \Pi(A) + \Pi(B) \tag{34}$$

where the condition $A \cap B = \emptyset$ turns out to be unnecessary. Zadeh's original concept, as well as "classical" (additive) probability measures, are viewed as the two limiting cases of this inequality (34).

But Giles ([13], p.189) then showed that any possibility measure defined by (33) can be written as: $\Pi(A) = \mathbf{Sup}_{P \in \mathcal{C}}(P(A))$ for a closed convex set $\mathcal{C}$ of probability measures P defined on a σ-algebra of $\mathcal{U}$, and *conversely*, that for any closed convex subset of probability measures (called *Belief Structure*), the functional defined by $\Pi(A) = \mathbf{Sup}_{P \in \mathcal{C}}(P(A))$ is a (generalized) possibility measure. Hence, it turns out that the generalized possibility theory of [13] is *exactly* the theory of upper-probabilities of paragraph 2.1.2, the dual necessity measures corresponding to lower-probabilities.

Inequality (18) can be written:

$$0 \leq P_X \leq \aleph \leq \Pi \leq T_X \leq 1$$

or equivalently, in DS notations : $0 \leq Bel \leq \aleph \leq \Pi \leq Pls \leq 1$ (35)

(Dubois and Prade obtained the same inequality in a constructive manner by approximating Belief and Plausibility functions by means of Possibility and Necessity measures, see [17]).

3.4 Fuzzy set theory

3.4.1 membership functions

Consider the restriction of the hitting functional T_X of a RACS X of $\mathcal{U}$ to the set $\mathcal{S}(\mathcal{U})$ of the singletons of $\mathcal{U}$. The relation $T_X(\{k\}) = \mathbf{Prob}(\{k\} \text{ hits } X) = \mathbf{Prob}(k \in X)$ suggests some similarity with the membership relation $\mu_S(x)$ of Fuzzy Set theory (S being a "fuzzy subset" of $\mathcal{U}$). Indeed, this analogy has been studied in the more general context of Random Sets ([18],[19]). Goodman [18] has established that *at least from a formal viewpoint, there exist systematic connections between fuzzy set theory and its operations, and probability theory and corresponding operations, via the concept of random sets.* Not only has it been established that fuzzy sets are nothing but *one-point coverages* of random sets, but all the classical "fuzzy operators" as well as their extensions, have a set-theoretic counterpart in the random set theory.

What about Random *Closed* sets (RACS)? It turns out that they correspond to *upper semi-continuous* fuzzy sets, i.e. fuzzy sets whose membership function is upper semi-continuous (u.s.c.) from $\mathcal{U}$ to $[0;1]$. Let us investigate this connection in more details.

Let μ_S be the membership function of a fuzzy subset S of a locally compact metrizable space $\mathcal{U}$. The set $\mu_S^{-1}([\xi;1])$ is called the *(closed) cross section of* S *at level* ξ and the mapping

$$\begin{aligned} \nu_S: \quad & [0;1] \longrightarrow \mathcal{P}(\mathcal{U}) \\ & \xi \longmapsto \nu_S(\xi) = \mu_S^{-1}([\xi;1]) \end{aligned} \tag{36}$$

is called the *(closed) cross section function* of S. It is well known ([6], pp.360-361) that μ_S is *upper semi-continuous* iff all the cross sections $\nu_S(\xi)$ are *closed* subsets of $\mathcal{U}$, i.e. iff ν_S takes its values in $\mathcal{F}(\mathcal{U})$ (instead of simply $\mathcal{P}(\mathcal{U})$, see figure 3).

Consider the relative topology of $\Re$ in $[0;1]$ and the Hit or Miss topology on $\mathcal{F}(\mathcal{U})$. Let us show that if μ_S is u.s.c. then the cross section function of S is *continuous* for these topologies. Since the Hit or Miss topology is generated by the basic opens $O'^K = \{F \in \mathcal{F}(\mathcal{U});\ F \text{ misses } K\}$ for K compact in $\mathcal{U}$, and $O'_O = \{F \in \mathcal{F}(\mathcal{U});\ F \text{ hits } O\}$ for O open in $\mathcal{U}$, it suffices to show that the inverse images by ν_S of these basic open sets are open for the relative topology in $[0;1]$.

$$\begin{aligned} \nu_S^{-1}(O'^K) &= \{\xi \in [0;1];\ \mu_S^{-1}([\xi;1]) \in O'^K\} \\ &= \{\xi \in [0;1];\ \mu_S^{-1}([\xi;1]) \text{ misses } K\} \\ &= \{\xi \in [0;1];\ \xi > \operatorname*{\mathbf{Sup}}_{x\in K} \mu_S(x)\} \\ &=]\operatorname*{\mathbf{Sup}}_{x\in K} \mu_S(x);1] \end{aligned} \qquad \begin{aligned} \nu_S^{-1}(O'_O) &= \{\xi \in [0;1];\ \mu_S^{-1}([\xi;1]) \in O'_O\} \\ &= \{\xi \in [0;1];\ \mu_S^{-1}([\xi;1]) \text{ hits } O\} \\ &= \{\xi \in [0;1];\ \xi < \operatorname*{\mathbf{Sup}}_{x\in O} \mu_S(x)\} \\ &= [0;\operatorname*{\mathbf{Sup}}_{x\in O} \mu_S(x)[\end{aligned}$$

and hence ν_S is continuous. Now, let ξ be a random variable valued in $[0;1]$, i.e. a measurable mapping from the probability space $(\Omega, \Sigma_\Omega, \mathbf{Prob})$ to the measurable space $([0;1], \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra of $[0;1]$. Since ν_S is continuous for our topologies, it is *measurable* for the Borel σ-algebras $\mathcal{B}$ and Σ'. The composition $X = \nu_S \circ \xi$ is a measurable function defined on $(\Omega, \Sigma_\Omega, \mathbf{Prob})$ and taking values in $(\mathcal{F}(\mathcal{U}), \Sigma')$ and is therefore a RACS. Its hitting capacity functional T_X is:

$$\begin{aligned} \forall K \in \mathcal{K},\quad T_X(K) &= \mathbf{Prob}(X^{-1}(F'_K)) = \mathbf{Prob}([\nu_S \circ \xi]^{-1}(\{F;\ F \text{ hits } K\})) \\ &= \mathbf{Prob}(\xi^{-1}(\nu_S^{-1}(\{F;\ F \text{ hits } K\}))) = \mathbf{Prob}(\xi^{-1}([0;\operatorname*{\mathbf{Sup}}_{x\in K} \mu_S(x)])) \\ &= \mathbf{Prob}(\xi \leq \operatorname*{\mathbf{Sup}}_{x\in K} \mu_S(x)) \end{aligned} \tag{37}$$

in particular, for $K = \{k\}$: $\quad T_X(\{k\}) = \mathbf{Prob}(\xi \leq \mu_S(k))$

If the random variable ξ is uniformly distributed over $[0;1]$, we have $\mathbf{Prob}(\xi \leq \alpha) = \alpha$, which gives $T_X(K) = \mathbf{Sup}_{k\in K}\, \mu_S(k)$ and in particular $T_X(\{k\}) = \mu_S(k)$. We call this RACS $X = \nu_S \circ \xi$ the *canonical RACS* associated with the u.s.c. fuzzy set S and we write $X = \kappa(S)$. X is a.s. non-empty ($T_X(\mathcal{U}) = 1$) iff μ_S is normalized in the sense of 3.3.1 ($\mu_S^{-1}(\{1\}) \neq \emptyset$).

Conversely, let X be a RACS of $\mathcal{U}$. The Choquet theorem insures that the capacity functional T_X of X is u.s.c. on $\mathcal{K}(\mathcal{U})$ equipped with the Hit or Miss topology. Obviously, T_X is also u.s.c. on the set $\mathcal{S}(\mathcal{U})$ of all singletons of $\mathcal{U}$, equipped with the induced Hit or Miss topology (remember that $\mathcal{S}(\mathcal{U}) \subset \mathcal{K}(\mathcal{U})$). But the canonical imbedding $\iota : \quad \mathcal{U} \longrightarrow \mathcal{S}(\mathcal{U})$ defined by $\iota(x) = \{x\}$ is a bicontinuous bijection (i.e. a homeomorphism). This can be checked quickly by noting that $\iota^{-1}(O'_O) = \iota^{-1}(\{\{x\} \in \mathcal{S}(\mathcal{U});\ \{x\} \text{ hits } O\}) = \iota^{-1}(\{\{x\};\ x \in O\}) = O$ and similarly that $\iota^{-1}(O'^K) = K^c$. Therefore, $\mu_S = T_X \circ \iota$ is u.s.c. for the topologies of $\mathcal{U}$ and $[0;1]$ and defines an u.s.c. fuzzy set S noted $S = \varphi(X)$. S is normalized iff X is a.s. non-empty.

We may sum up the above considerations by saying that every RACS defines a (unique) u.s.c. fuzzy set which is its point interpretation (called "point coverage" in [18]), and for every u.s.c. fuzzy set, there exists a (not necessarily unique) RACS of which it is a point interpretation. One such RACS, called canonical, is constructed by uniformly randomizing the cross sections of the fuzzy set membership function.

3.4.2 T-norms/conorms and Fuzzy connectives

Fuzzy connectives are binary operators used for combining two fuzzy sets defined over the same Universe $\mathcal{U}$. Triangular norms and conorms (resp. T-norms and T-conorms) are binary operators that verify the basic axioms of commutativity, associativity and monotonicity as well as specific boundary conditions. It has been argued that the Fuzzy connectives *intersection* and *union* should obey this axiomatics if they are to be compatible with the intuitive concepts of *conjunction* and *disjunction*.

It is well known that the **Min** operator is the greatest T-norm $\wedge$ whereas **Max** is the smallest T-conorm $\vee$:

$$\forall (a,b) \in [0;1]^2, \quad \begin{array}{c} \wedge(a;b) \leq \mathbf{Min}(a;b) \\ \mathbf{Max}(a;b) \leq \vee(a;b) \end{array}$$

Several families of T-norms and conorms have been suggested, including: $\wedge(a,b) = \mathbf{Min}\left(1;(a^p+b^p)^{1/p}\right)$ and the corresponding conorm: $\vee(a,b) = \mathbf{Max}\left(0;1-((1-a)^p+(1-b)^p)^{1/p}\right)$ (where $p \in [1;+\infty]$ is a parameter). For this family of T-norms/conorms, we obviously have:

$$\forall (a,b) \in [0;1]^2, \quad \begin{array}{c} \mathbf{Max}(0;a+b-1) \leq \wedge(a;b) \leq \mathbf{Min}(a;b) \\ \mathbf{Max}(a;b) \leq \vee(a;b) \leq \mathbf{Min}(1;a+b) \end{array} \tag{38}$$

which reminds us of:

$$\mathbf{Max}(0;T_{X_1}(\{x\})+T_{X_2}(\{x\})-1) \leq T_{X_1 \cap X_2}(\{x\}) \leq \mathbf{Min}(T_{X_1}(\{x\});T_{X_2}(\{x\})) \tag{11}$$

$$\mathbf{Max}(T_{X_1};T_{X_2}) \leq T_{X_1 \cup X_2} \leq \mathbf{Min}(1;T_{X_1}+T_{X_2}) \tag{8}$$

THEOREM 1. *For any T-norm $\wedge$ verifying* (38) *and any pair* (S_1, S_2) *of semi-continuous fuzzy subsets of* $\mathcal{U}$ *, there exists a pair* (X_1, X_2) *of RACS such that* S_1 *(resp.* S_2*) is a point interpretation of* X_1 *(resp.* X_2*) and* $S_1 \wedge S_2$ *is a point interpretation of* $X_1 \cap X_2$*:* $S_1 = \varphi(X_1)$, $S_2 = \varphi(X_2)$ *and* $S_1 \wedge S_2 = \varphi(X_1 \cap X_2)$.

This follows from the remarks of paragraph 2.5, and is a consequence of the Choquet theorem. A slightly stronger formulation is:

THEOREM 1′. *For any T-norm $\wedge$ verifying* (38) *and any pair* (S_1, S_2) *of semi-continuous fuzzy subsets of* $\mathcal{U}$ *, there exists a RACS* X_2 *such that* S_2 *is a point interpretation of* X_2 *and* $S_1 \wedge S_2$ *is a point interpretation of* $\kappa(S_1) \cap X_2$*:* $S_2 = \varphi(X_2)$ *and* $S_1 \wedge S_2 = \varphi(\kappa(S_1) \cap X_2)$.

Conversely:

THEOREM 2. *The binary operator $\wedge$ defined on the set of (semi-continuous) fuzzy subsets of $\mathcal{U}$ by:* $S_1 \wedge S_2 = \varphi(\kappa(S_1) \cap \kappa(S_2))$ *is commutative, associative and verifies the boundary conditions of a T-norm, but is not necessarily monotonic. Several operators can be obtained, some of which are T-norms, depending on the choice of the statistical dependence between the 2 RACS* $\kappa(S_1)$ *and* $\kappa(S_2)$.

Commutativity and associativity of $\wedge$ result from that of $\cap$ (as a RACS operator) which themselves result from that of $\cap$ (as a set-theoretic operator). $0 \wedge 0 = 0$ and $1 \wedge \alpha = \alpha$ result from (38). To show that $\wedge$ is not necessarily monotonic, consider two fuzzy sets S_1 and S_2 such that $0 < \mu_{S_1}(x) \leq \mu_{S_1}(y)$ and $\mu_{S_2}(x) \leq \mu_{S_2}(y)$ at some points $x, y \in \mathcal{U}$ and such that $\mathbf{Prob}(y \in \kappa(S_2) \mid y \in \kappa(S_1)) = 0$ and $\mathbf{Prob}(x \in \kappa(S_2) \mid x \in \kappa(S_1)) > 0$. Then obviously: $\mu_{S_1}(y) \wedge \mu_{S_2}(y) = T_{\kappa(S_1) \cap \kappa(S_2)}(\{y\}) = \mathbf{Prob}(y \in \kappa(S_1)) \cdot \mathbf{Prob}(y \in \kappa(S_2) \mid y \in \kappa(S_1)) = 0 < \mathbf{Prob}(x \in \kappa(S_1)) \cdot \mathbf{Prob}(x \in \kappa(S_2) \mid x \in \kappa(S_1)) = T_{\kappa(S_1) \cap \kappa(S_2)}(\{x\}) = \mu_{S_1}(x) \wedge \mu_{S_2}(x)$ which violates monotonicity.

Now if $\kappa(S_1)$ and $\kappa(S_2)$ are taken statistically independent, $\mu_{S_1}(x) \wedge \mu_{S_2}(x) = T_{\kappa(S_1) \cap \kappa(S_2)}(\{x\}) = T_{\kappa(S_1)}(\{x\}) \cdot T_{\kappa(S_2)}(\{x\}) = \mu_{S_1}(x) \cdot \mu_{S_2}(x)$ and $\wedge$ reduces to a simple product, which is a T-norm verifying (38).

The corresponding results for T-conorms are obtained by changing $\wedge$ into $\vee$ and $\cap$ into $\cup$.

3.5 summary of the mathematical connections

The connections between the various formalisms presented in this section, when $\mathcal{U}$ is a compact metrizable topological space, are summarized in figures 4 and 5.

4. Discussion: what is the point of comparing two theories?

The purpose of section 3 was to establish the mathematical connections between various theories that have a similar goal: the representation of imprecise, uncertain or fuzzy knowledge in the context of Artificial Intelligence. The comparison was based on the axiomatics of the theories, i.e. we considered two theories to be (mathematically) *equivalent* whenever their respective sets of axioms can be shown to imply each other. Note that it has long been argued that axiomatic definitions of physical theories are highly desirable (see [20] for a recent discussion).

Are such comparisons necessary, or even simply useful? After all, some authors have argued that even though two theories are shown to be mathematically equivalent, they should not be confused as long as their *interpretations* differ.

4.1 Of the necessity of theoretical comparisons: axiomatics vs interpretation

From a philosophical point of view, the need for establishing theoretical links between different theories that have similar goals and contexts is undoubtedly a matter of scientific conscientiousness. Isolating a theory and preventing external criticism or comparison with other scientific theories does not do any good to its long term reputation, although it does play a protective role for its growth in the early stages ([21]). The Dempster-Shafer and the Fuzzy set theories provide excellent illustrations in this respect. Both have been intentionally isolated ever since their creations (1976 for the former [1], 1965 for the latter) and a few attempts to compare them with other formalisms ([5],[12],[18]) or with each other ([22]) have been largely ignored in spite of their important practical implications.

A strong and recent defense of the DS theory can be found in [16]. The author thwarts any criticism by arguing that even though two theories can be shown to be mathematically related, they should not be compared on the basis of their axiomatics since their *interpretations* may differ: *"That both models share the same mathematical properties is not an argument for them being the same concept. Remember that water flow and electricity can be described mathematically by the same differential equations - but water is not electricity"* (from [16]).

At the risk of being considered retrograde, we strongly question the idea that current computers can do more than merely process information in compliance with a given mathematical framework (be it arithmetics, predicate logics or whatever). This leaves us with only two alternatives: either we define the "interpretation" of a theory in terms of another "meta" framework, or we rely on a human "expert" to carry out the final interpretation.

If the latter is chosen, how can a computer system distinguish between two mathematically equivalent theories?

Indeed, water <u>is</u> electricity for a computer as long as both are modelled by the same mathematical entities: one needs a "meta-knowledge" (such as the common-sense of human beings) to distinguish between them.

If the former definition is selected, the mathematical links between the two "sub-theories" actually make them interchangeable for the meta-framework. Suppose that a computer is provided with the extra knowledge that water has mass but electricity does not. It is then able to distinguish between the two concepts but it does not hurt to assume that water and electricity are identical as long as mass is not involved.

Similarly, we cannot see any reason why a computer could distinguish between the DS formalism and the theory of a.s. non-empty RACS of a discrete finite space, unless a meta-framework explicitely differentiates the two (by providing a formal definition of belief that has no counterpart in RACS theory, for instance). In spite of what many authors suggest, Dempster's rule of combination does NOT provide this meta-framework, as it is a mere conditional RACS intersection.

The remarkable work of [21] is a successful attempt at recasting the foundations of the Fuzzy set theory by getting rid of the ontological assumption about the "fuzziness of the world". Indeed, the claim that "the World is inherently fuzzy" has served as the main justification of the Fuzzy set theory, and provided a very efficient protection against external criticism/comparison. [21] provides a solid foundational basis for the comparison of the Fuzzy set formalism with other scientific theories.

4.2 Of the usefulness of theoretical comparisons

Mere scientific conscienciousness does not make computers more intelligent, nor does it make programing them easier. This paragraph intends to show how theoretical comparisons can be useful in practice, as they create a basis for the extension of theories that suffer from limited scopes and they allow for hybrid techniques that exploit the mathematical links between them.

As a matter of fact, it often happens that a seemingly promising theory, with lots of "new ideas", fails to fulfill its potential due to some limitations in its formulation. These limitations may be fundamental assumptions on which the whole theory is constructed, but they often are technical requirements added for the sake of simplicity and to facilitate the emergence of the new concepts in a simplified context.

If some mathematical connections have been established with another formalism, it may be possible to rid the new theory of its technical limitations by using these links as a basis for a *theoretical extension*. The example given below

(5.1) illustrates this idea: the original DS theory (as found in [1]) is limited to *finite* Universes $\mathcal{U}$ and an extension to general compact metrizable (topological) spaces is proposed, based on the equivalence of the original formalism with the theory of a.s. non-empty Random Closed Sets of a discrete finite Universe. This extension allows for finite, countably infinite as well as uncountably infinite spaces.

The above mentioned equivalence could also be used in another direction to extend the original DS formalism to general (not necessarily a.s. non-empty) Random Closed sets of a discrete finite Universe. This is equivalent to dropping conditions (19-i) and (20-ii) in the definition of Belief and Plausibility functions. It turns out that this extension has already been suggested elsewhere ([23],[24],[16]) without recourse to the RACS theory. Note that [23] uses the links between DS theory and another mathematical formalism (1st order predicate calculus) to derive the extension.

Another undesirable limitation of the DS formalism is the requirement that all pieces of evidence be independent from each other. A consequence of this limitation is the incompatibility of the concept of *idempotence*: it is not possible to tell whether Dempster's rule is idempotent or not since we cannot combine the same information twice. We believe that idempotence is an important concept in AI and the RACS based extension of Dempster's rule allows for the representation of idempotence.

Another useful consequence of theoretical comparisons is the ability of building *hybrid techniques* that take advantage of the theoretical connections between several theories so that the tools of one of them can be used in the other or in conjunction with the tools of the other. This is especially useful when one of the theories is weak at dealing with some aspects of the problem. For example, it has been generally acknowledged that the absence of a systematic scheme to construct Belief/Plausibility functions and Fuzzy set membership functions from World evidence or expert information is a weak point of both the DS and the Fuzzy set theories. We argue (par. 5.3) that RACS theory is well-equipped to deal with such difficulties.

5. Exploiting the mathematical connections

5.1 extension of Dempster-Shafer theory to compact metrizable spaces

The finiteness assumption is definitely a serious limitation of the DS theory; consider the classical stereo-vision problem, for instance: although all images are digital and represented on a finite grid, the results of applying a stereo-vision algorithm on a pair of such images are truly 3-dimensional features. If we want to represent this information on a finite Frame of Discernment, we must sample this continuous representation space too!

Hence, it appears desirable to *extend* the DS formalism in some suitable way, so as to be able to use finite, countably infinite, as well as uncountably infinite (continuum) Frames. How is this to be done?

A straightforward way would be to transpose directly to the most general space the definitions (19) and (20) for Belief and Plausibility functions, hence keeping the intuitive interpretation of Belief Intervals in terms of lower and upper probabilities. As a matter of fact, this approach is indeed possible and the resulting extended DS theory turns out to be equivalent to the theory of *a.s. non-empty Random Sets* (not necessarily closed, see [5]).

For the reasons given in paragraph 2.2 however, we do not consider this straightforward extension as the most suitable for practical and experimental purposes and we argue that it should take place within a topological setting. We shall thus extend the DS theory to infinite topological spaces on the ground of equivalence (3.2.1) by identifying it with the theory of a.s. non-empty RACS. As a matter of fact, this identification is entirely straightforward if we restrict it to *compact* 2nd countable (Hausdorff) spaces since we saw that these spaces many desirable properties. In particular, these properties insure that the inclusion capacity functional P_X defined by (3) entirely determines the RACS X and since we shall identify P_X with a Belief function Bel, this property is required.

Concluding the preceding constructive steps, *we define the Extended DS theory as the theory of almost surely non-empty RACS of a compact metrizable Universe.* Note that it has been argued that *compact spaces are natural generalizations of finite spaces* ([25]), and in view of figure 1, we could even say that *compact 2nd countable spaces are natural extensions of finite discrete spaces* so that it appears natural to consider the extension of the DS theory to such spaces...

The only delicate point in this extension is related to the Mass functions (22): in general, there is no reason why the probability distribution P' of a RACS X should have a *density* $\frac{dP'}{d\nu}$ with respect to some measure ν on $\mathcal{F}(\mathcal{U})$. If this happens to be the case, then the Mass function is identified with this density and we may apply general integral formulas to generalize equations (23).

As for Dempster's rule $\oplus$, it can be generalized through equation (27), and thus becomes valid in *all cases*: $\mathcal{U}$ finite or infinite, countable or uncountable, X_1 and X_2 independent or not. This rule is *idempotent* because RACS intersection is (which is a consequence of the idempotence of $\cap$ considered as a set operator). Remember that the idempotence requirement only constrains the combination of RACS that are identical *as mappings* (and not only *in probability*). Thus idempotence cannot be made explicit in (29), or in Shafer's original work [1], where all pieces of evidence are assumed independent.

Although conflict was presented by Shafer as a useful measure of the "disagreement" between several pieces of evidence, engineers have been increasingly aware of its painful and somewhat cumbersome behaviour. Imagine that we succeeded in combining 100 complex pieces of evidence in a large Frame $\mathcal{U}$ (at a non-negligeable cost!), and that we are presented with an extra (101st) piece of evidence. What a disappointment if it happens to flatly conflict with the other 100! What should we do then? Modify the Frame of Discernment $\mathcal{U}$ as suggested by Shafer himself? But then, the whole accumulation process looses its order-independence.

In an attempt to get rid of conflict, one of the authors suggested ([23]) *augmenting* the Frame of Discernment by adding to $\mathcal{U}$ an extra element ∇ called the *hidden element.* He used the 1st order predicate calculus to show that every Frame of Discernment is in fact associated with a *fundamental assumption*, which he made explicit in the Frame's *Characteristic Formula.* Since the *negation* of this fundamental assumption is a valid (logical) proposition, it must be incorporated in the Frame itself and the focal elements of any evidence must be of the form $\{\nabla\} \cup A \subset \bar{\mathcal{U}} = \{\nabla\} \cup \mathcal{U}$. Since such focal elements cannot be disjoint (they intersect at least at $\{\nabla\}$), there is no conflict and the normalization constant in (30) disappears ($\kappa = 0$).

A very similar result was obtained by Smets [24], who argued that when one constructs a Frame of Discernment $\mathcal{U}$, one must choose between the *Closed World Assumption* (CWA) and the *Open World Assumption* (OWA). The former explicitly states that $\mathcal{U}$ encompasses all the possible evidence ("there is no unknown evidence"), while the latter more reasonably postulates the existence of a set $\mathcal{U}^*$ of "unknown evidence". Of course, nothing is known about this set, but its very existence (i.e. $\mathcal{U}^* \neq \emptyset$) gets rid of the normalization factor in (30). [24] concludes that the DS theory as given in [1] implicitely makes the Closed World Assumption, hence generating conflict.

These results can also be obtained by topological considerations. We know that a locally compact 2nd countable Hausdorff space $\mathcal{U}$ can always be compactified into a compact metrizable space $\bar{\mathcal{U}}$ by adding a single *point at infinity* ∞. A non-empty RACS $\bar{X}$ of $\bar{\mathcal{U}}$ is simply a random variable taking values in the compact 2nd countable (Hausdorff) space $\bar{\mathcal{F}} \setminus \{\emptyset\} = \big(\mathcal{K}(\mathcal{U}) \setminus \{\emptyset\}\big) \cup \big\{F \cup \{\infty\};\ F \in \mathcal{F}(\mathcal{U})\big\}$. We can identify 3 cases α, β, γ for an a.s. non empty RACS $\bar{X}$ of $\bar{\mathcal{U}}$:

α) $T_{\bar{X}}(\{\infty\}) = 0$. $\bar{X}$ is an a.s. non-empty compact RACS of $\mathcal{U}$. Dempster's rule is not always possible. We are in the situation of the Closed World Assumption (CWA).

β) $T_{\bar{X}}(\{\infty\}) > 0$. Such Random Closed Sets are always combinable by Dempster's rule since they all hit with non-zero probability at least a common element. In a word, they cannot be "totally conflicting".

γ) $T_{\bar{X}}(\{\infty\}) = 1$. $\bar{X}$ can almost surely be written as $\bar{X} = \{\infty\} \cup X$, where X is a RACS of $\mathcal{U}$. Since this is a subcase of the previous case, such Random Closed Sets are always combinable by Dempster's rule. In fact, their relative conflict is always 0 since they all hit (a.s.) a common element. We are in the situation of the Open World Assumption (OWA).

Clearly, the "Augmented Frame of Discernment" of [23] is equivalent to the one-point compactification of the Universe $\mathcal{U}$ and we saw that the latter exists iff $\mathcal{U}$ is locally compact Hausdorff (and not merely completely regular). Since this is one of the basic assumptions of the RACS theory, the approach of [23] is fully justified in our extended DS formalism based on the RACS theory. In other words, the RACS-based extension of the DS framework entitles us to reduce all the "unknown evidence" of $\mathcal{U}^*$ to a single ∇ element.

Note that as Dempster's rule is based on a set-theoretic operator (intersection), it may be suitable for problems involving imprecision only, but not for problems where both imprecision *and* uncertainty are important factors. For such problems, a probabilistic operator (2.5) should be used.

5.2 construction of Fuzzy sets from Belief/Plausibility functions and vice versa

Consider general Belief/Plausibility functions Bel and Pls of a compact metrizable topological Universe $\mathcal{U}$. Is it possible to construct a (u.s.c.) Fuzzy set from Bel or Pls? If so, how is this to be done?

Given the interpretation of Belief and Plausibility functions in terms of the including and hitting capacity functionals of a RACS (3.2.2) and the links between Fuzzy sets and RACS (3.4.2 and figure 5), the answer to both questions is simple: $Pls \circ \iota$ defines the membership function of a Fuzzy set and we showed (3.4.1) that it is upper semi-continuous, hence defining a u.s.c. Fuzzy set of $\mathcal{U}$. The membership value of a point of $\mathcal{U}$ is its plausibility (or communality).

Conversely, for a given upper semi-continuous Fuzzy set S with membership function μ_S,

$$\forall A \in \mathcal{U}, \qquad \begin{aligned} Pls(A) &= \operatorname*{Sup}_{x \in A} \mu_S(x) \\ Bel(A) &= \operatorname*{Inf}_{x \in A^c} (1 - \mu_S(x)) \end{aligned}$$

define Plausibility and Belief functions, canonically obtained from S and related to the Hitting and Including capacity functionals of the canonical RACS $\kappa(S)$ associated with S. It turns out that these functions are also Possibility and Necessity measures with distribution μ_S whenever μ_S is normalized (cf. 3.3.1). Therefore, we can say that normalized u.s.c. Fuzzy sets canonically define Zadeh Possibility/Necessity measures (which are particular Plausibility/Belief functions), but it is of course possible to define "non-canonical" Belief and Plausibility functions from a normalized u.s.c. Fuzzy set, and these will *not* be Possibility/Necessity measures. One way of doing this is by randomizing the cross section of μ_S using a non-uniform random variable ξ: if f is the probability density of ξ and $F(x) = \int_0^x f(y)dy$ its cumulative probability distribution, then

$$\forall A \in \mathcal{U}, \quad \begin{aligned} Pls(A) &= F\big(\operatorname*{Sup}_{x\in A} \mu_S(x)\big) \\ Bel(A) &= F\big(\operatorname*{Inf}_{x\in A^c}(1-\mu_S(x))\big) \end{aligned}$$

are Plausibility and Belief functions which are Possibility and Necessity measures *if and only if* f is uniform (constant).

5.3 construction of Belief/Plausibility functions from World evidence

One of the weak points of the DS formalism is the absence of any systematic scheme for constructing Belief/Plausibility functions from pieces of evidence. The following theorem particularizes one such construction scheme:

THEOREM 3. *The only order-independent, piecewise and point-compatible combination operator that allows the construction of general Belief/Plausibility functions from (closed) subsets of a Universe is the Mean operator.*

A proof is given in the appendix, in the equivalent RACS framework. "order-independent" stands for "commutative and associative" and by "general Belief/Plausibility functions", we mean non-trivial (0-1 valued) Belief/Plausibility functions.

A *point-compatible combination* is such that the Belief function constructed from subsets of $\mathcal{U}$ reduces to an ordinary point probability measure (Bayesian Belief function) whenever all subsets reduce to singletons. This requirement guarantees the compatibility with the point approach, or in other words, insures that our construction scheme is an extension of the point approach. It can also be viewed as an *economy principle*: if all pieces of evidence are points of $\mathcal{U}$, there is no need for a set theory and since ordinary point probabilities are sufficient, they should be used.

A *piecewise combination* is such that the combined Belief (or Plausibility) of a proposition (or subset of the Universe) A depends only on A, and not on any other proposition $B \neq A$.

The order-independence requirement is rather natural and hardly questionable. To better see what a non-piecewise combination would look like, imagine two experts providing some knowledge about propositions (subsets) of a (compact metrizable) Universe $\mathcal{U}$ in terms of Belief/Plausibility/Communality functions (Bel_i, Pls_i, q_i for $i = 1, 2$). We want to combine the knowledge provided by both experts and determine the Belief (Bel), Plausibility (Pls) or

Communality (q) functions associated with the result of this combination. The "piecewise combination" requirement insures that, for every proposition $A \in \mathcal{U}$, $Bel(A)$, $Pls(A)$ and $q(A)$ depend only on A. Since everything we know about proposition A is the knowledge provided by the two experts ($Bel_i(A)$, $Pls_i(A)$ and $q_i(A)$ for $i = 1, 2$), we conclude that $Bel(A)$, $Pls(A)$ and $q(A)$ must be functions of $Bel_i(A)$, $Pls_i(A)$ and $q_i(A)$ ($i = 1, 2$) and of these values only: $Bel(A) = f(Bel_1(A), Bel_2(A); Pls_1(A), Pls_2(A); q_1(A), q_2(A))$, etc. If we drop the requirement, then $Bel(A)$ may depend on $Bel(B)$ or $Pls(B)$ or $q(B)$ for another proposition $B \neq A$. In the extreme case, $Bel(A)$ could depend on the Belief values of *all* the propositions of $\mathcal{U}$, in which case we say that combination is a *global operation.*

Finally, THEOREM 3 only constrains the construction of Belief/Plausibility functions from *subsets* of $\mathcal{U}$. Indeed, (closed) subsets of $\mathcal{U}$ are the most basic pieces of evidence one can obtain from the World (they are binary in nature), and in fact, they are the only ones that are *directly accessible* by simple *physical* measurements from the World, without involving complex cognitive processes.

Hence, if one wants to use an order-independent and point-compatible operator different from the Mean probabilistic operator, one must drop at least one of the following premises:

1) combination is not piecewise: the combined Belief of a proposition may depend on the Belief of other propositions.

2) the construction of Belief functions is piecewise but does not deal directly with subsets of $\mathcal{U}$: the inputs are other Belief functions. We must assume that "Belief functions can be found in the World" and measured directly without any construction process from subsets of $\mathcal{U}$. As this is obviously not the case in *physical World*, it is clear that we are working in a *human World*, where human beings (experts, witnesses, etc) are the only sources of information.

In systems where the sources of information are not human, or at least *not only* human, we must exclude possibility 2 above: the construction of Belief must be either probabilistic (Mean operator) or global (non-piecewise).

5.4 construction of Fuzzy sets from World evidence

It is clear that the Fuzzy set theory suffers from the same weakness as the DS formalism, namely the absence of any systematic scheme for constructing Fuzzy set membership functions from World evidence. Some authors even argue that this absence prevents it from being a "scientific theory", relagating it to a mere "engineering technique"...

In view of the relation between Plausibility/Communality functions and (u.s.c.) Fuzzy membership functions, it is clear that the following corollary of THEOREM 3 holds:

THEOREM 3′. *The only order-independent, point-compatible and piecewise combination operator that allows the construction of general (u.s.c.) Fuzzy sets from (closed) subsets of a Universe is the Mean operator.*

The strongest argument against one of the premises of the above theorem is the ontological claim that "both the physical and human Worlds are fuzzy" according to which fuzzy membership functions can be measured directly even in the physical World and used for further combinations. This claim has been discussed at length in [21] and seems difficult to refute for the human World. However, engineers have been increasingly aware of the difficulty of measuring membership functions in the physical World experimentally without using some sort of statistical construction. And indeed, there are some good reasons to think that the simple pieces of information "belongs/does not belong to a set" precede the more complex membership functions, and thus that the latter can be induced from the former.

If we reject the "fuzzy physical World" claim, we may conclude from THEOREM 3′: in systems where the sources of information are not human, or at least *not only* human, the construction of Fuzzy membership functions must be either probabilistic (Mean operator) or global (non-piecewise). Note that this rejects the classical "fuzzy connectives" **Min** and **Max**! Indeed these operators are order-independent and piecewise, but do not allow the construction of general (=non-crisp) u.s.c. Fuzzy sets from subsets of the Universe.

6. Concluding remarks

Mathematically related theories may differ in their interpretations, but unless these are part of a meta-framework, they are entirely arbitrary. Some authors ([1], [24]) have strongly argued that the DS formalism should not be confused with any other, and in particular that it is "non-probabilistic" in nature (cf. introduction of [24]). But as Dubois and Prade point out in a comment of [16] (p.282), *Shafer has reinterpreted Dempster's upper and lower probabilities in terms of personal plausibility and belief. However, he has just modified the terminology.* Terminologies are important since they link pure mathematics to physical or conceptual entities, but they are *interpretations* too and therefore arbitrary. Indeed, computers are rather indifferent to terminologies and no computer will ever be able to distinguish between the allegedly "non-probabilistic" DS formalism and the theory of a.s. non-empty Random

(Closed) Sets.

Considering these equivalences, why did we choose the RACS approach instead of one of its equivalent or related forms? The 3 components (imprecision, uncertainty, topology) of this approach play the role of *axioms* of the resulting theory, but they can be (at least partially) justified: Stone representation theorem for Set theory, betting behaviour and "scoring rules"([3]) for Probability theory and Mathematical Morphology ([7],[8],[26]) for topology. RACS theory is general enough for most practical applications, including those that involve uncountably infinite spaces, but not so general as to deal with useless, purely mathematical abstractions. The concept of statistical dependence makes it richer than the DS formalism and as a matter of fact, intuitive properties, such as idempotence, cannot be expressed in the latter.

The RACS approach, therefore, provides a powerful and sufficient framework for the representation of imprecision, uncertainty and fuzziness in Artificial Intelligence, as well as a unified view of the concepts of belief/plausibility, possibility/necessity, upper/lower probabilities and semi-continuous fuzzy sets.

Appendix

Proof of PROPOSITION 1:

- (24) $\Longrightarrow$ (25)

Suppose that f verifies (24) and let $A'_1, \dots, A'_n$ be any finite family of subsets of $\mathcal{U}$. Then let $A_0 = \bigcap_{i=1,\dots,n} A'_i$ and $A_i = A'_i, \quad i = 1, \dots, n$. $\{A_0, A_1, \dots, A_n\}$ is a finite family of subsets of $\mathcal{U}$ and we can apply (24): $f(A_0) = f(\bigcap_{i=1,\dots,n} A'_i) \leq \sum_{\substack{I \subset \{1,\dots,n\} \\ I \neq \emptyset}} (-1)^{|I|+1} f(\bigcap_{j=1,\dots,n} A'_j \cup \bigcup_{i \in I} A'_i)$ and since $\bigcap_{j=1,\dots,n} A'_j \subset \bigcup_{i \in I} A'_i$, equation (25) follows. Furthermore, if $A \subset B$ then $f(A) \leq f(A \cup (B \setminus A)) = f(B)$ and f is increasing.

- (25) $\Longrightarrow$ (24)

Suppose that f is increasing and verifies (25). Let $A_0, A_1, \dots, A_n$ be any finite family of subsets of $\mathcal{U}$ and $A'_i = A_i \cup A_0, \quad i = 1, \dots, n$. Then $A_0 \subset A_0 \cup \bigcap_{i=1,\dots,n} A_i = \bigcap_{i=1,\dots,n} (A_0 \cup A_i) = \bigcap_{i=1,\dots,n} A'_i$ And since f is increasing, we can write: $f(A_0) \leq f(\bigcap_{i=1,\dots,n} A'_i) \leq \sum_{\substack{I \subset \{1,\dots,n\} \\ I \neq \emptyset}} (-1)^{|I|+1} f(\bigcup_{i \in I} A'_i)$ (the rightmost inequality resulting from (25)) and since $\bigcup_{i \in I} A'_i = \bigcup_{i \in I} (A_0 \cup A_i) = A_0 \cup \bigcup_{i \in I} A_i$, we obtain (24). QED

A dual proposition can be obtained by changing f into $1 - f^c$.

PROPOSITION 2:

Let $(\Omega, \Sigma_\Omega, \mathbf{Prob})$ *be a probability space,* $\mathcal{U}$ *be a compact metrizable space and* $\Sigma_\mathcal{U}$ *its Borel* σ*-algebra . Let* $\mathcal{C}$ *be a closed convex set of (regular,* σ*-additive) probability measures. Then there exists a* $(\Sigma_\Omega, \Sigma_\mathcal{U})$*-strongly measurable mapping* Γ *defined on* Ω *and taking values in* $\mathcal{F}(\mathcal{U}) \subset \mathcal{P}(\mathcal{U})$ *such that:*

$$\forall A \in \Sigma_\mathcal{U}, \quad \underset{P\in\mathcal{C}}{\mathbf{Inf}}\, P(A) = \mathbf{Prob}(\Gamma \subset A) \quad \text{and} \quad \underset{P\in\mathcal{C}}{\mathbf{Sup}}\, P(A) = \mathbf{Prob}(\Gamma \text{ hits } A) \tag{39}$$

if and only if *class* $\mathcal{C}$ *verifies:*

$$\forall n \geq 1, \quad \forall (A_1, \ldots, A_n) \in (\Sigma_\mathcal{U})^n, \quad \forall P \in \mathcal{C} \qquad P(\bigcup_{i=1}^{n} A_i) \geq \sum_{I \subset \{1,\ldots,n\}} (-1)^{|I|+1}\, \underset{P\in\mathcal{C}}{\mathbf{Inf}}(P(\bigcap_{i\in I} A_i)) \tag{40}$$

or equivalently:

$$\forall n \geq 1, \quad \forall (A_1, \ldots, A_n) \in (\Sigma_\mathcal{U})^n, \quad \forall P \in \mathcal{C} \qquad P(\bigcap_{i=1}^{n} A_i) \leq \sum_{I \subset \{1,\ldots,n\}} (-1)^{|I|+1}\, \underset{P\in\mathcal{C}}{\mathbf{Sup}}(P(\bigcup_{i\in I} A_i)) \tag{40'}$$

Note that (39) implies $\mathbf{Prob}(\Gamma = \emptyset) = 0$, i.e. Γ must be a.s. non-empty.

• Let us first show the "only if" part of the proposition. The $(\Sigma_\Omega, \Sigma_\mathcal{U})$-strong measurability of mapping $\Gamma : \Omega \Longrightarrow \mathcal{F}(\mathcal{U})$ implies its (Σ_Ω, Σ') measurability, where $\Sigma' = \Sigma_{\mathcal{F}(\mathcal{U})}$ is the Borel σ-algebra of $\mathcal{F}(\mathcal{U})$ equipped with the Hit or Miss topology induced by the topology of $\mathcal{U}$. Hence, Γ is a RACS of $\mathcal{U}$ and its capacity functional P_Γ (resp. T_Γ) must verify the conditions of Choquet's theorem ((1) and (3) resp.). But $P_\Gamma(A) = \mathbf{Prob}(\Gamma \subset A) = \underset{P\in\mathcal{C}}{\mathbf{Inf}}\, P(A)$ and $T_\Gamma(A) = \mathbf{Prob}(\Gamma \text{ hits } A) = \underset{P\in\mathcal{C}}{\mathbf{Sup}}(P(A))$, so that $\underset{P\in\mathcal{C}}{\mathbf{Inf}}$ and $\underset{P\in\mathcal{C}}{\mathbf{Sup}}$ must statisfy (3-iii) and (1-iii), respectively. Since both functionals are increasing, Propositions 1 and 1′ hold and they must verify (25′) and (25) respectively, which obviously imply (40) and (40′).

• Conversely, let $P_* = \underset{P\in\mathcal{C}}{\mathbf{Inf}}$ and $P^* = \underset{P\in\mathcal{C}}{\mathbf{Sup}}$. Conditions (i) and (ii) of the Choquet theorem (3) (resp. (1)) are verified by P_* (resp. P^*) by construction, and condition (3-iv) (resp. (1-iv)) results from Proposition 1′ (resp. 1).

The *inner regularity* of the measures $P \in \mathcal{C}$ implies (see e.g. [27], p.448): $\forall P \in \mathcal{C}, \quad P(O_n) \uparrow P(O)$ whenever $O_n \uparrow O$ in $\mathcal{O}$. This obviously implies: $\underset{P\in\mathcal{C}}{\mathbf{Sup}}(P(O_n)) \uparrow \underset{P\in\mathcal{C}}{\mathbf{Sup}}(P(O))$ whenever $O_n \uparrow O$, i.e. $P^* = \underset{P\in\mathcal{C}}{\mathbf{Sup}}$ is l.s.c. on $\mathcal{O}$, which implies that P^* is u.s.c. on $\mathcal{K}$ ([4], p.31).

Similarly, the *outer regularity* of the measures $P \in \mathcal{C}$ implies: $\forall P \in \mathcal{C}, \quad P(K_n) \downarrow P(K)$ whenever $K_n \downarrow K$ in $\mathcal{K}$. This obviously implies: $\underset{P\in\mathcal{C}}{\mathbf{Inf}}(P(K_n)) \downarrow \underset{P\in\mathcal{C}}{\mathbf{Inf}}(P(K))$ whenever $K_n \downarrow K$, i.e. $P_* = \underset{P\in\mathcal{C}}{\mathbf{Inf}}$ is u.s.c. on $\mathcal{K}$ or equivalently l.s.c. on $\mathcal{O}$ (condition (3-iii)).

By the Choquet theorem, P_* (resp. P^*) is therefore the capacity functional P_X (resp. T_X) of a (a.s. non-empty) RACS X, which is nothing but a multivalued mapping strongly measurable w.r.t. Σ_X and $\Sigma_{\mathcal{U}}$. QED

Proof of PROPOSITION 3:

If the proposition holds for possibility measures, it holds for necessity measures by duality. Let Π be a Zadeh possibility measure. Conditions (31-i and ii) obviously imply (20-i and ii). Let us show by recursion that (20-iii) is also satisfied:

- for $n = 2$, (20-iii) is $\forall (A, B) \in \mathcal{P}(\mathcal{U})^2$, $\Pi(A \cap B) \leq \Pi(A) + \Pi(B) - \Pi(A \cup B)$. But $\Pi(A \cup B) = \text{Max}(\Pi(A), \Pi(B))$: if $\Pi(A) \geq \Pi(B)$, then $\Pi(A \cup B) = \Pi(A)$ and since Π is increasing $\Pi(A \cap B) \leq \Pi(B)$. By exchanging A and B, (20-iii) holds if $\Pi(A) \leq \Pi(B)$ too.
- let $n > 2$ be an integer and suppose that (20-iii) holds for n. Let $(A_1, \ldots, A_n, A_{n+1}) \in \mathcal{P}(\mathcal{U})^{n+1}$. We can write:

$$\begin{aligned}
\sum_{I \subset \{1,\ldots,n+1\}} (-1)^{|I|+1} \underset{i \in I}{\text{Max}}\{\Pi(A_i)\} = & \sum_{I \subset \{2,\ldots,n\}} (-1)^{|I|+1} \underset{i \in I}{\text{Max}}\{\Pi(A_i)\} \\
& + \sum_{I \subset \{2,\ldots,n\}} (-1)^{|I|+1} \text{Max}\{\Pi(A_1), \Pi(A_{n+1}), \underset{i \in I}{\text{Max}}\{\Pi(A_i)\}\} \\
& - \sum_{I \subset \{2,\ldots,n\}} (-1)^{|I|+1} \text{Max}\{\Pi(A_1), \underset{i \in I}{\text{Max}}\{\Pi(A_i)\}\} \\
& - \sum_{I \subset \{2,\ldots,n\}} (-1)^{|I|+1} \text{Max}\{\Pi(A_{n+1}), \underset{i \in I}{\text{Max}}\{\Pi(A_i)\}\}
\end{aligned} \tag{41}$$

Suppose that $\Pi(A_{n+1}) \geq \Pi(A_1)$. Then $\text{Max}\{\Pi(A_1), \Pi(A_{n+1}), \underset{i \in I}{\text{Max}}\{\Pi(A_i)\}\} = \text{Max}\{\Pi(A_{n+1}), \underset{i \in I}{\text{Max}}\{\Pi(A_i)\}\}$ and (41) reduces to

$$\begin{aligned}
\sum_{I \subset \{1,\ldots,n+1\}} (-1)^{|I|+1} \underset{i \in I}{\text{Max}}\{\Pi(A_i)\} = \sum_{I \subset \{1,\ldots,n\}} (-1)^{|I|+1} \underset{i \in I}{\text{Max}}\{\Pi(A_i)\} & \geq \Pi\Big(\bigcap_{i=1,\ldots,n} A_i\Big) \quad \text{(recursion assumption)} \\
& \geq \Pi\Big(A_{n+1} \cap \bigcap_{i=1,\ldots,n} A_i\Big) \quad (\Pi \text{ increasing})
\end{aligned}$$

Similarly, if $\Pi(A_{n+1}) < \Pi(A_1)$, $\text{Max}\{\Pi(A_1), \Pi(A_{n+1}), \underset{i \in I}{\text{Max}}\{\Pi(A_i)\}\} = \text{Max}\{\Pi(A_1), \underset{i \in I}{\text{Max}}\{\Pi(A_i)\}\}$ and thus:

$$\begin{aligned}
\sum_{I \subset \{1,\ldots,n+1\}} (-1)^{|I|+1} \underset{i \in I}{\text{Max}}\{\Pi(A_i)\} = \sum_{I \subset \{2,\ldots,n+1\}} (-1)^{|I|+1} \underset{i \in I}{\text{Max}}\{\Pi(A_i)\} & \geq \Pi\Big(\bigcap_{i=2,\ldots,n+1} A_i\Big) \quad \text{(recursion assumption)} \\
& \geq \Pi\Big(A_1 \cap \bigcap_{i=2,\ldots,n+1} A_i\Big) \quad (\Pi \text{ increasing})
\end{aligned}$$

which proves (20-iii) for $n+1$. QED

Proof of THEOREM 3:

Let us prove the theorem in the equivalent RACS framework.

Let $X_1, ..., X_n$ be n subsets of the (compact metrizable) Universe $\mathcal{U}$, $T_{X_1}, ..., T_{X_n}$ (resp. $P_{X_1}, ..., P_{X_n}$ and $R_{X_1}, ..., R_{X_n}$) their hitting (resp. including and implying) functionals. Our goal is to construct a RACS X from the n (deterministic) RACS X_i using an order-independent piecewise and point-compatible operator. This is equivalent (Choquet theorem) to constructing capacity functionals T_X and P_X from X_i.

The piecewise requirement insures that, for every compact $K \in \mathcal{K}(\mathcal{U})$, $T_X(K)$ and $P_X(K)$ depend only on K. In a general topological space, the only information one can get from the two given sets is whether they *hit or miss* each other or whether one is *included* in the other. And indeed, the only information we have concerning K is contained in $T_{X_1}(K), \ldots, T_{X_n}(K)$, $P_{X_1}(K), \ldots, P_{X_n}(K)$ and $R_{X_1}(K), \ldots, R_{X_n}(K)$, which are all binary values (since X_i are deterministic). Hence, $T_X(K)$ and $P_X(K)$ should be functions of only these values:

$$T_X(K) = f(T_{X_1}(K), \ldots, T_{X_n}(K); P_{X_1}(K), \ldots, P_{X_n}(K); R_{X_1}(K), \ldots, R_{X_n}(K))$$

$$P_X(K) = g(T_{X_1}(K), \ldots, T_{X_n}(K); P_{X_1}(K), \ldots, P_{X_n}(K); R_{X_1}(K), \ldots, R_{X_n}(K))$$

Let K_1 and K_2 be two compacts such that $X_1 \subset K_1 \cup K_2$, $X_1 \not\subset K_1$, $X_1 \not\subset K_2$ and X_1 hits $K_1 \cap K_2$, and such that the other X_i $(i \neq 1)$ have the same spatial relations with K_1 as with K_2, $K_1 \cup K_2$ and $K_1 \cap K_2$ (we can always find $X_1, \ldots, X_n$ such that such K_1 and K_2 exist). Since, in general:

$$T_X(K_1) \leq T_X(K_1 \cup K_2) \leq T_X(K_1) + T_X(K_2) - T_X(K_1 \cap K_2)$$

we get:

$$\begin{aligned} &f(1, ..., T_{X_n}(K_1); 0, ..., P_{X_n}(K_1); R_{X_1}(K_1), ..., R_{X_n}(K_1)) \\ &\quad \leq f(1, ..., T_{X_n}(K_1); 1, ..., P_{X_n}(K_1); R_{X_1}(K_1), ..., R_{X_n}(K_1)) \\ &\quad \leq f(1, ..., T_{X_n}(K_1); 0, ..., P_{X_n}(K_1); R_{X_1}(K_1), ..., R_{X_n}(K_1)) \end{aligned}$$

from which we deduce:

$$\begin{aligned} &f(1, ..., T_{X_n}(K_1); 0, ..., P_{X_n}(K_1); R_{X_1}(K_1), ..., R_{X_n}(K_1)) \\ &\quad = f(1, ..., T_{X_n}(K_1); 1, ..., P_{X_n}(K_1); R_{X_1}(K_1), ..., R_{X_n}(K_1)) \end{aligned}$$

which shows that f does not depend on P_{X_1}. Similarly, f does not depend on P_{X_i} $(i = 1, \ldots, n)$:

$$T_X(K) = f(T_{X_1}(K), \ldots, T_{X_n}(K); R_{X_1}(K), \ldots, R_{X_n}(K))$$

By considering two compacts such that $K_1 \cap K_2 \subset X_1$, $K_1 \not\subset X_1$ and $K_2 \not\subset X_1$, and such that the other X_i $(i \neq 1)$ have the same spatial relations with K_1 as with K_2, $K_1 \cup K_2$ and $K_1 \cap K_2$, we can show that g does not depend on R_{X_i}:

$$P_X(K) = g(T_{X_1}(K), \ldots, T_{X_n}(K); P_{X_1}(K), \ldots, P_{X_n}(K))$$

Now reduce all pieces of evidence to singletons of $\mathcal{U}$: $X_i = \{x_i\}$. Then, $T_{X_i}(K) = P_{X_i}(K)$ for all k, and thus:

$$P_X(K) = g'(T_{X_1}(K), \ldots, T_{X_n}(K))$$

Our point-compatibility requirement insures that the resulting RACS X should be a random *point* variable and we know that this is so iff $P_X = T_X$. Hence, we must have:

$$f(T_{X_1}(K), \ldots, T_{X_n}(K); R_{X_1}(K), \ldots, R_{X_n}(K)) = g'(T_{X_1}(K), \ldots, T_{X_n}(K))$$

which proves that f cannot depend on R_{X_i} and $f = g'$. Now, the order-independence assumption insures that f only depends on the total *number* of 1s and 0s and not their positions as arguments. This is the same as saying that f depends only on the *sum* of its (binary) arguments: $f(\alpha_1, \ldots, \alpha_n) = f^*(\sum_{i=1}^n \alpha_i)$. Since X is a random point, T_X must be additive, which implies that f^* itself must be additive: $\forall p \in \{0, \ldots, n-1\}, \quad f^*(p+1) = f^*(p) + f^*(1)$. Noticing that $T_X(\mathcal{U}) = 1 = f(1, \ldots, 1) = f^*(n)$ and $T_X(\emptyset) = 0 = f(0, \ldots, 0) = f^*(0)$, we conclude: $\forall p \in \{0, \ldots, n\}, \quad f^*(p) = p/n$ and hence:

$$\forall K \in \mathcal{K}(\mathcal{U}), \qquad T_X(K) = \frac{\sum_{i=1}^n T_{X_i}(K)}{n}$$

QED

References

[1] G. Shafer, *A mathematical Theory of Evidence*, Princeton University Press, 1976.

[2] P. Cheeseman, "In Defense of Probability" *Proc. Int. J. Conf. Artif. Intell.*, vol. IJCAI85, pp.1002-1009, 1985.

[3] D.V. Lindley, "Scoring Rules and the Inevitability of Probability" *Int. Stat. Rev.*, vol. 50, pp.1-26, 1982.

[4] G. Matheron, *Random Sets and Integral Geometry*, Wiley, New York, 1975.

[5] H.T Nguyen, "On Random Sets and Belief Functions" *J. of Math. Anal. and Appl.*, vol. 65, pp.531-542, 1978.

[6] N. Bourbaki, "General Topology" in *Elements of Mathematics*, Hermann, Paris, 1966.

[7] J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press, 1982.

[8] J. Serra, "The Boolean Model and Random Sets" in *Image Modeling*, Academic Press (A..Rosenfeld ed.), 1981.

[9] N. Cressi.e. and G.M. Laslett, "Random Set Theory and Problems of Modeling" *SIAM Review*, vol. 29, Dec. 1987.

[10] D.G. Kendall, "Foundations of a Theory of Random Sets" in *Stochastic Geometry*, Wiley (Harding&Kendall eds), 1974.

[11] A. Dempster, "Upper and Lower Probabilities Induced by a Multivalued Mapping" *Ann. of Math. Stat.*, vol. 38, pp.325-339, 1967.

[12] H.E. Kyburg, "Bayesian and Non-Bayesian Evidential Updating" *Artif. Intell.*, vol. 31, pp.271-293, 1987.

[13] R. Giles, "Foundations for a Theory of Possibility" in *Fuzzy Information and Decision Processes*, North-Holland (Gupta&Sanchez eds), 1982.

[14] T. Garvey, "Evidential Reasoning for Geographic Evaluation for Helicopter Route Planning" *IEEE Trans. Geoscience and Remote Sensing*, vol. GE25, pp.294-304, May 1987.

[15] L.A. Zadeh, "Fuzzy Sets as a basis for a theory of possibility" *Fuzzy Sets and Syst.*, vol. 1, pp.3-28, 1978.

[16] Ph. Smets, "Belief Functions" in *Non standard logics for automated reasoning* (Smets, Mandani, Dubois and Prade eds), Academic Press, London, pp.253-286, 1988 (also in *IEEE Pattern Anal. Machine Intell.* vol. PAMI12, May 1990).

[17] D. Dubois and H. Prade, "On several Representations of an Uncertain Body of Evidence" in *Fuzzy Information and Decision Proceses*, North-Holland (Gupta&Sanchez eds), 1982.

[18] I.R. Goodman, "Characterizations of N-ary Fuzzy Set Operations which induce Homomorphic Random Set operations" in *Fuzzy Information and Decision Processes*, North-Holland (Gupta&Sanchez eds), 1982.

[19] Wang Pei-Zhuang and E. Sanchez, "Treating a Fuzzy Subset as a projectable Random Subset" in *Fuzzy Information and Decision Processes*, North-Holland (Gupta&Sanchez eds), 1982.

[20] G. Ludwig, "An Axiomatic Basis as a desired Form of a Physical Theory" in *Logic, Methodology and Philosophy of Science VIII*, Elsevier Science Publishers (Fenstad et al. eds), pp.447-457, 1989.

[21] G. Tamburrini and S. Termini, "Some Foundational Problems in the Formalization of Vagueness" in *Fuzzy Information and Decision Processes*, North-Holland (Gupta&Sanchez eds), 1982.

[22] J. Kampé de Fériet, "Interpretation of Membership functions of Fuzzy Set in terms of Plausibility and Belief" in *Fuzzy Information and Decision Processes*, North-Holland (Gupta&Sanchez eds), 1982.

[23] T. Matsuyama, "On logical Foundations of Evidential Reasoning based on the Dempster-Shafer Probabilistic Model" *J. of the Japanese Soc. for Artif. Intell.*, vol. 4, pp.104-114, May 1989 (in japanese).

[24] Ph. Smets, "Transferable Belief Model versus Bayesian Model" *Proc. Europ. Conf. on Artif. Intell.*, vol. ECAI88, pp. 495-500, August 1988.

[25] E. Hewitt, "The rôle of Compactness in Analysis" *Amer. Math. Month.* vol. 67, pp.499-516, June/July 1960.

[26] J. Serra, *Image Analysis and Mathematical Morphology*, vol. 2, Academic Press, 1988.

[27] M.M. Rao, *Measure Theory and Integration*, Wiley & Sons, 1987.

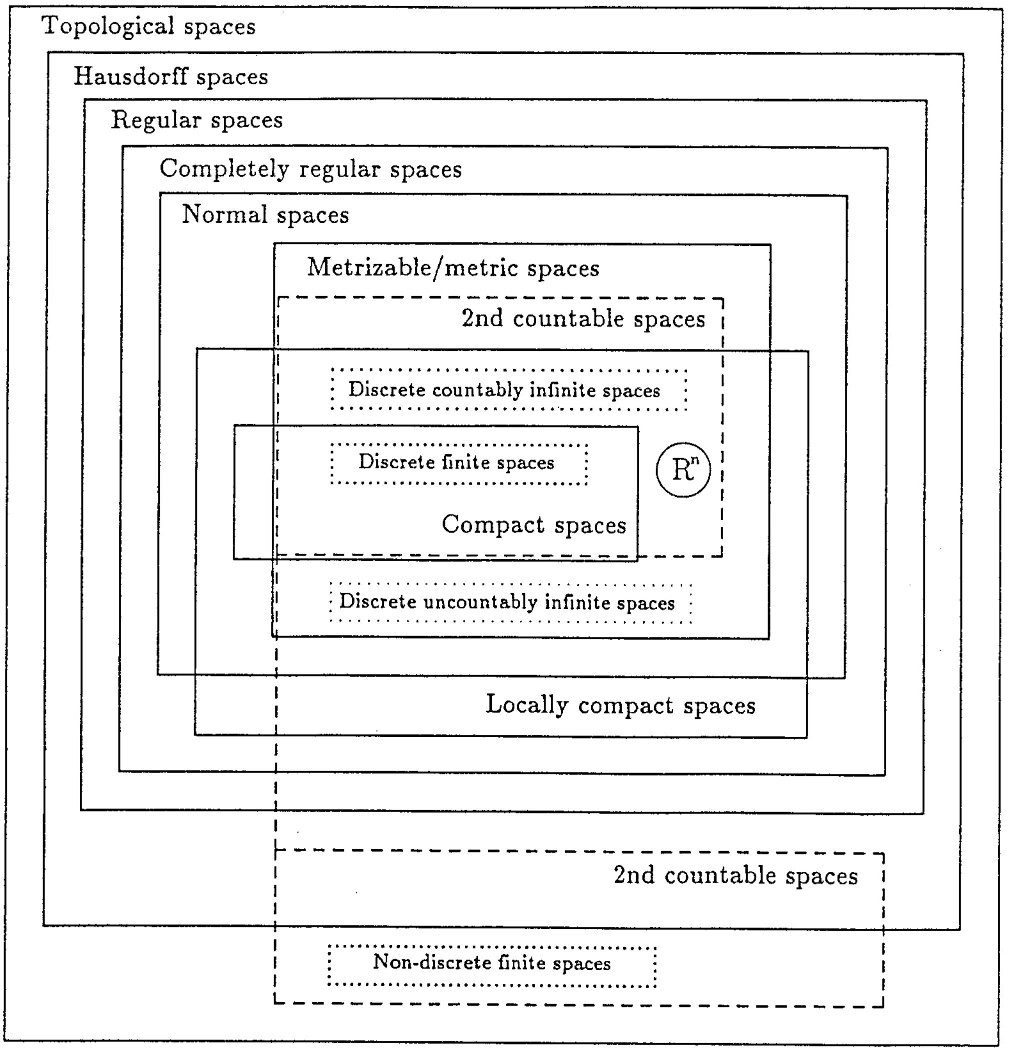

Fig.1 The main types of topological spaces

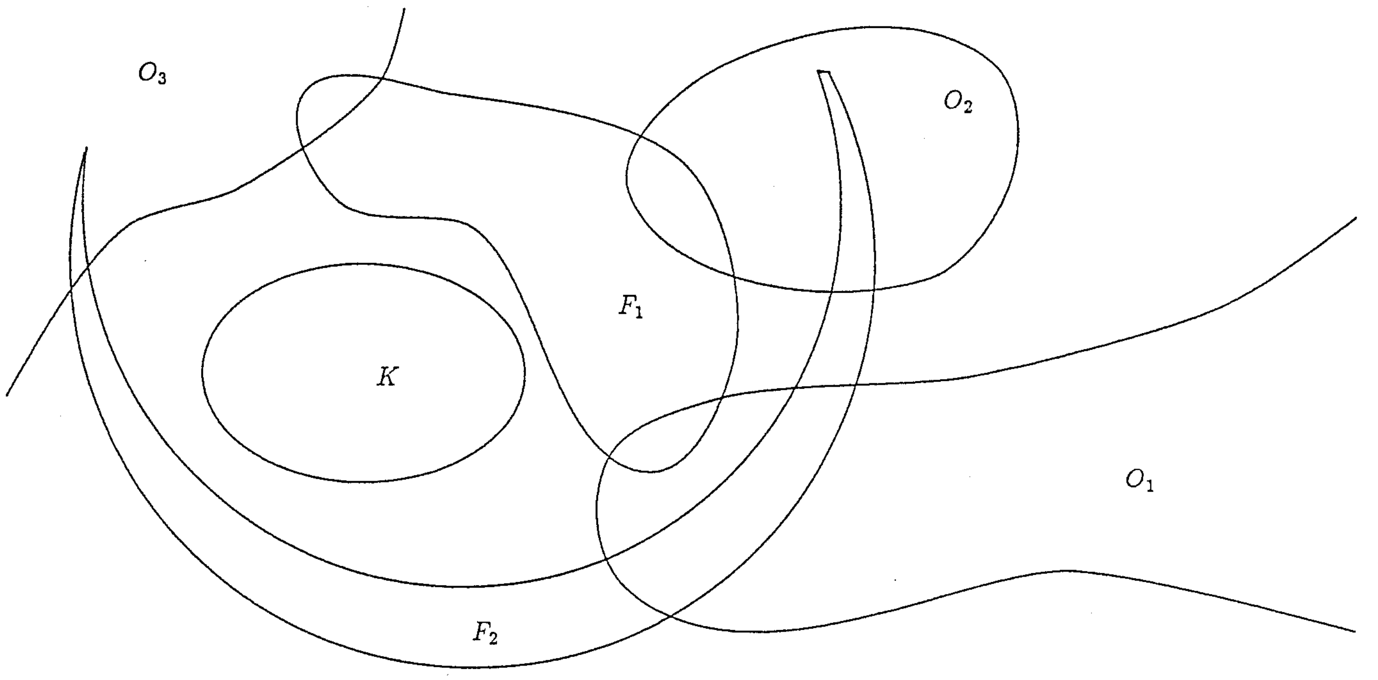

Fig.2: Hit or Miss topology (F_1 and F_2 belong to the same neighborhood $O'_{O_1} \cap O'_{O_2} \cap O'_{O_3} \cap O'^{K}$)

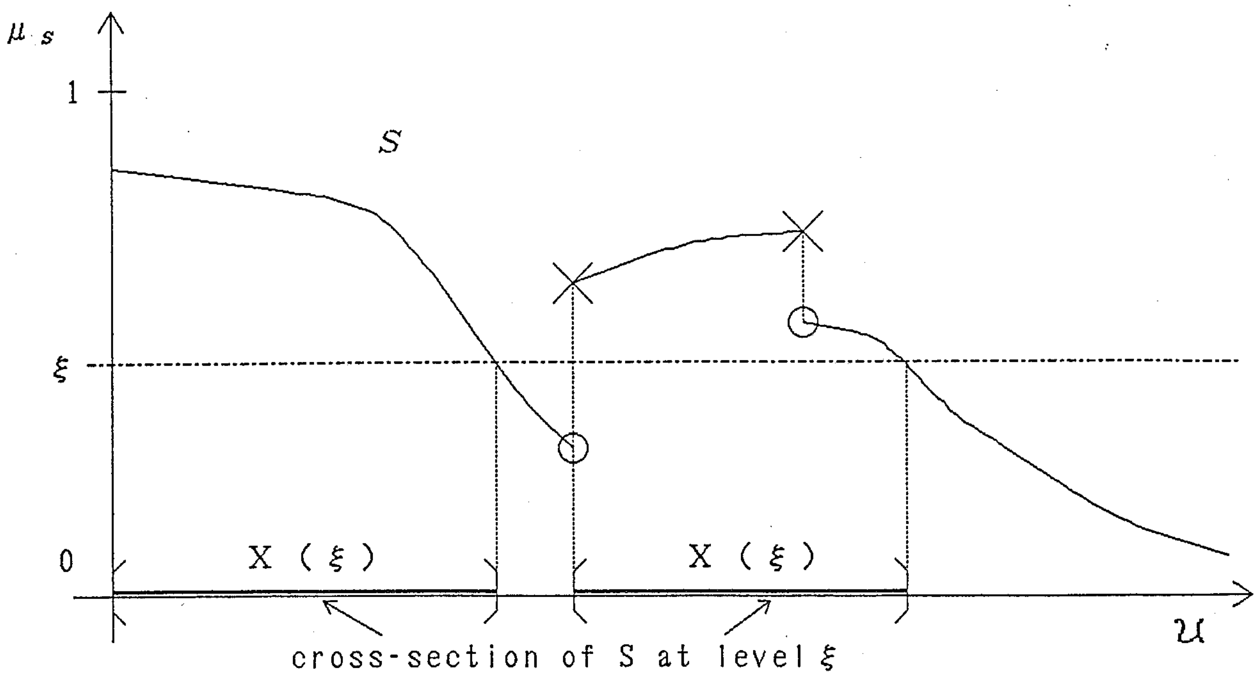

Fig.3: Cross section of an u.s.c. fuzzy set

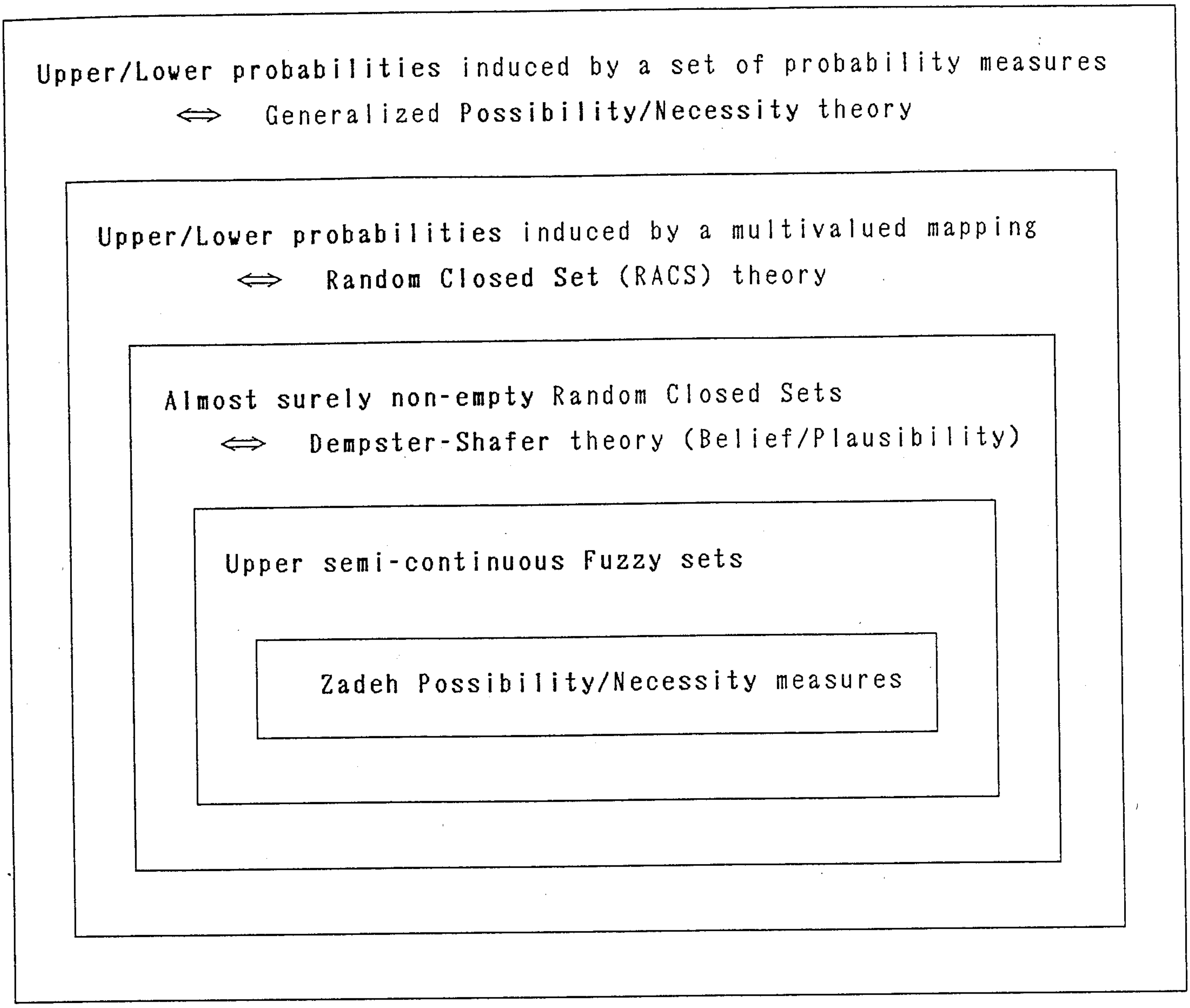

Fig.4: Mathematical links between the theories of section 3 in a compact metrizable topological space

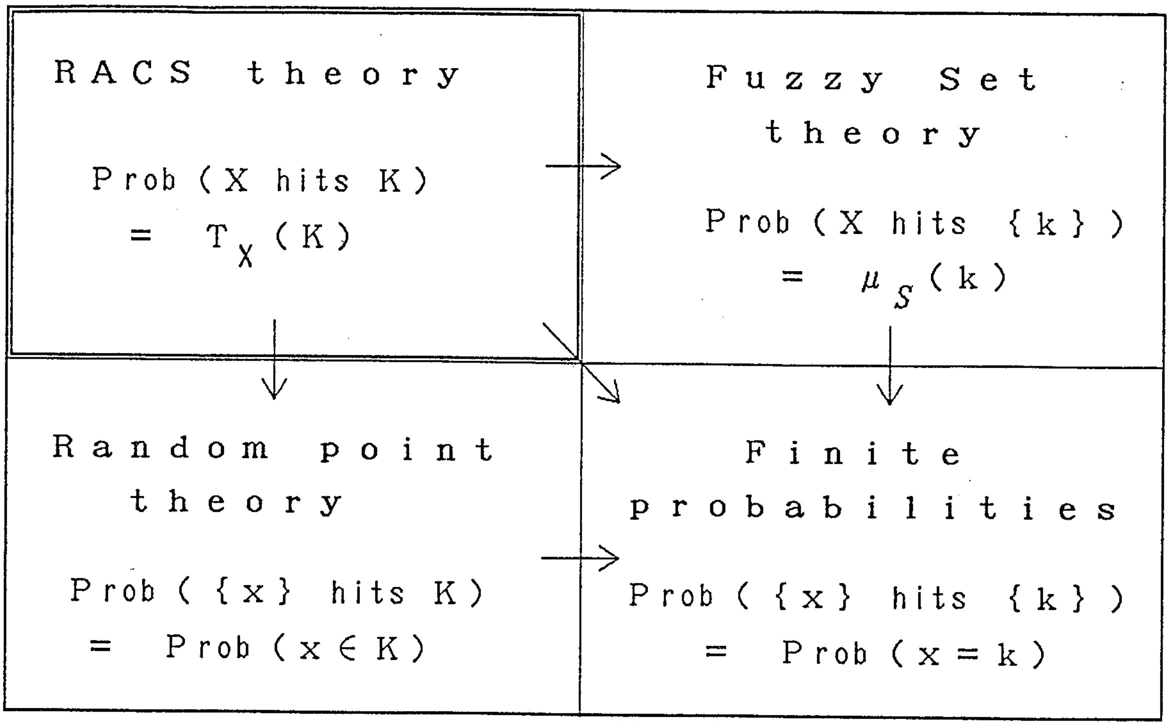

Fig.5: Both the classical Random point theory and the theory of (u.s.c.) Fuzzy sets are particularizations of the RACS theory, as singletons are particular compact subsets.