# What's in a Fuzzy Set?

**Marco Piastra**

Dipartimento di Informatica e Sistemistica
Università degli Studi di Pavia
Via Ferrata, 1, I-27100 Pavia, Italy
pmarco@vision.unipv.it

**Abstract**

A modified version of the first-order logic of probability presented in (Halpern 1990) – with probability on possible worlds – makes it possible to formulate an alternative characterisation of fuzzy sets. In this approach, fuzzy sets are no longer seen as primitive entities with an intuitive justification, but rather as structured entities emerging in a suitable logical framework. Some fuzzy techniques of practical relevance are shown to be encodable in this way. In addition, the resulting approach leads to a clearer epistemological analysis in that it clarifies the purposive nature of the kind of uncertainty that can be modelled by fuzziness.

## 1. Introduction

A fairly common characterisation of fuzzy sets in the literature is as a primitive notion – usually but not necessarily related with uncertainty – which is given intuitive justification in terms of similarity with classical sets. Typically, a textbook on the topic will start by describing the real-valued generalisation of classical characteristic functions and will then discuss a few examples of fuzzy set modelling applied to everyday knowledge immediately after.

From an epistemological standpoint, this attitude would appear to assume an incremental modelling strategy for the notion itself. One disadvantage, deriving from the rather weak premises, is that further formal and informal elements have to be introduced at all subsequent stages of theoretical development. For instance, a viable mathematical definition for the fundamental set-theoretic operators – i.e. conjunction, disjunction and complement – is typically achieved by positing an ensemble of natural axioms (e.g. continuity, commutativity, monotonicity, etc.). However sensible, these axioms are restricted to each group of items to be defined and are grounded on pragmatic intuition alone. Not surprisingly, as the formal apparatus grows, the method becomes less and less effective.

The problem is particularly noticeable with the closely-related field of fuzzy logic. Paris (1994) for instance gives a set of natural axioms for negation, conjunction and disjunction and then skates over the detailed assessment of an implication-like connective, as "it seems far less clear what axioms should hold for this function". Clearly, the situation does not improve when defining a fuzzy-logical consequence relation.

The central claim of this paper is that fuzzy sets could be characterised in a different way, namely as a derivative notion in a suitable logical framework. More precisely, fuzzy sets may be made to emerge from a logical 'deep structure' which governs their 'surface' behaviour. Although the modal probabilistic framework that has to be adopted is somewhat elaborate, the resulting characterisation leads to clearer epistemological analysis. In addition, the approach would appear to preserve the qualitative flavour of the original setting. At the very least, it can be shown that some well-known fuzzy techniques can be encoded in the proposed framework.

The paper first describes the modal probabilistic logical framework and subsequently provides a tentative definition of fuzzy sets with respect to this framework. The possible encoding for existing fuzzy techniques is then discussed in some detail. In so doing, the informal notion of fuzziness is reconsidered again and the relationship with Gärdenfors' notion of *conceptual spaces* (1992) is discussed in detail.

The ideas presented here owe much to the influence of a number of other studies. As already stated, the logical apparatus is derived from (Halpern 1990). The construction of a possible world semantics for fuzzy logic was first explored in (Ruspini 1991), whereas the characterisation of fuzzy sets given here has some similarity with that developed by Gerla (1994), who instead adopts probability on formulas, and in (Wang and Tan 1997). To a large extent, the treatment of compositionality follows the line presented in (Dubois and Prade 1994).

## 2. A Modal Probabilistic Framework

The logical framework we are seeking in this section has the two main goals of providing:

- a language for expressing statements about fuzzy sets;
- a clearly-defined consequence relation.

Sometimes, fuzzy logic is investigated by assuming a many-valued propositional logic as a basis and by establishing a relation with fuzzy set theory. Here we follow a different path. We shall start by constructing a logic of fuzzy sets, whereas the relationship with many-valued logics will be discussed in a subsequent section. Note that the achievement of an axiomatisation is not assumed as a goal in this context. One significant reason is that, as we shall see below, this goal cannot be achieved at all. Another reason is that most fuzzy logic techniques are applied in practice as numerical techniques. Hence the informal goal in this paper is not to develop a form of automated reasoning as a replacement, but rather a suitable formal encoding that allows us to assess the soundness of the techniques in question.

As mentioned above, the construction of the framework is based on the first order logic of probability on possible worlds developed by Halpern (1990). At the outset, we have a first order language $\Phi$ with equality containing predicate and function symbols of various arities, together with a family of *object constants* and *object variables*. Terms of this first sort are to be interpreted, as usual, in a domain of objects $D$. It is further assumed that the language $\Phi$ includes a name for every object in $D$.

In order to simplify the exposition that follows, we adopt a modal translation of Halpern's original language following the approach described in (Voorbraak 1993). This translation makes use of the two special modal symbols $\Box$ and $\Box_p$, where $p$ is real number in [1, 0]. Informally, the meaning of a formula $\Box\varphi$ is that $\varphi$ is true in every possible world (see below), whereas $\Box_p\varphi$ means that $\varphi$ is true with a probability at least equal to $p$.

Define a probability structure $M$ as a tuple $(D, W, \pi, S, \mu)$, where $D$ is a domain of objects, $W$ is a set of possible worlds, $\pi$ is a binary function assigning the proper meaning in each world $w \in W$ to the symbols in $\Phi$. Define also a valuation function $v$ assigning an element of $D$ to each *object term* in $\Phi$. To keep things manageable, we assume that $v$ does not vary across $W$, i.e. that all terms are *rigid*. Note however that this does not say anything about the extensions of predicate symbols, which are in fact allowed to vary. Semantic rules for $(M, v, w) \models \varphi$ are defined as usual in first order logic, i.e. by induction over the structure of $\varphi$, with the sole exception that a reference to the world $w$ is kept as required by the function $\pi$. The function $\mu$ is a countably additive probability measure over the algebra of sets $\{w : (M, w) \models \varphi\}$ generated by the sentences $\varphi$ in $\Phi$. The set $S$ is defined as follows:

$$S = \{w : \forall\varphi,\ \mu(\varphi) = 1 \Rightarrow (M, w) \models \varphi\}$$

in other words, $S$ is the subset containing *random worlds* (see Gaifman and Snir 1982); i.e. worlds that do not satisfy any sentence with a zero probability of being true. The semantic rules for the two modal symbols are:

- $(M, v) \models \Box\varphi$ iff $\forall\, w \in S,\ w \models \varphi$
- $(M, v) \models \Box_p\varphi$ iff $\mu(\{w \in W : w \models \varphi\}) \geq p$

which yield the expected meaning. Note that both definitions are given w.r.t. to $S$ and this validates the identity:

$$\Box\varphi \equiv \Box_1\varphi.$$

We also introduce two derivative modal symbols $\Diamond_p$ and $\boldsymbol{P}_p$ defined as follows:

$$\Diamond_p\varphi \equiv \neg\Box_{1-p}\neg\varphi$$
$$\boldsymbol{P}_p\varphi \equiv (\Box_p\varphi \wedge \neg\Diamond_p\varphi).$$

The modal operator $\boldsymbol{P}_p$ expresses a point-valued probability constraint. Note that, syntactically, modal operators might be nested. Voorbraak (1993) proves that in a setting of this kind – which is of modality KD45 – any nested formula has a non-nested equivalent.

We assume that any formula $\phi$ in the final language of the framework is a composition:

$$\lambda : \psi$$

where $\psi$ is a formula in the first order language with modal extensions. The *label* $\lambda$ contains symbols from the first order language $\Phi$ plus symbols from a language of a second sort, namely the language of real closed fields. The latter includes the binary predicates > and =, the binary functions + and $\times$, the three constants –1, 0 and 1 plus *field functions* and *field variables*. The interpretation of this second sort of symbols is given as usual, i.e. with respect to the set of real numbers $\mathbf{R}$. In addition, $\lambda$ may also contain *measuring functions* (Bacchus 1990), i.e. functions mapping object terms to field terms. Finally, we allow *field variables* to appear in $\psi$ as indexes to the modal operators $\Box_p$. As a convention, we shall use the letters $x$, $y$, $z$ for object variables and $p$, $q$, $r$, for field variables.

Informally, the label $\lambda$ acts as a *generalised quantifier* binding both sorts of variables in $\psi$. From a formal standpoint, however, the labelled notation is a mere notational facility for preserving a clear separation between the algebraic part and the logical part of each formula. The semantics of labelled formulas is given in terms of a translation rule. Let:

$$Q_1 v_1 \ldots Q_n v_n\, \lambda^*$$

be the prenex form of $\lambda$, where each $Q_i v_i$ represents a quantifier $Q_i$ applied to a variable $v_i$ of either kind. Hence, by definition, a labelled formula $\lambda : \psi$ is a shorthand notation for:

$$\forall v_{n+1} \ldots \forall v_m\ Q_1 v_1 \ldots Q_n v_n\, (\lambda^* \to \psi)$$

where $v_{n+1} \ldots v_m$ are variables of either kind occurring free in $\lambda^*$. The semantic rule for labelled formulas $\phi = \lambda : \psi$ is defined with respect to a class of extended structures $M'$ where $v'$ is an extended valuation function assigning a real value to field terms as well. We write $M' \models \phi$ iff for every valuation $v'$, $(M', v') \models \phi$. The consequence relation is defined in the usual way – i.e. given a set of labelled formulas $\Sigma$, we write $\Sigma \models \phi$ iff every structure $M'$ satisfying the formulas in the set $\Sigma$ also satisfies $\phi$.

A first example of a labelled formula is:

$$p\ :\ \forall x\, (\Box_p A(x) \to \Box_p B(x)).$$

By convention, the term $p$ in the label $\lambda$ is taken here as the abbreviation of $p = p$, which is satisfied by any valuation. Hence in the above formula $p$ is universally quantified; the formula states that, for any object in $D$, the probability of its being $B$ is at least equal to that of its being $A$. A more interesting example of labelled notation relates to the definition of conditional formulas:

$$(q = 0 \wedge p = 0) \vee (q \geq 0 \wedge r = p \times q)\ :$$
$$(\boldsymbol{P}_q\beta \wedge \boldsymbol{P}_r(\alpha \wedge \beta)) \to \boldsymbol{P}_p(\alpha \mid \beta)$$

which is the axiom given in (Bacchus 1990). A ternary modality of conditional independence can be defined as:

$p = q \times r$ :

$$(\alpha, \beta \perp \gamma) \equiv ((\boldsymbol{P}_q(\alpha \mid \gamma) \wedge \boldsymbol{P}_r(\beta \mid \gamma)) \rightarrow \boldsymbol{P}_p(\alpha \wedge \beta \mid \gamma)).$$

The analysis of finitary properties, however, reveals that the framework is intractable. One of the proofs contained in (Halpern 1990) can be adapted to show that any formula in his setting, with probability on possible worlds, can be translated into a labelled formula of the above kind. Hence the proofs contained in (Abadi and Halpern 1991) also demonstrate that the framework presented is hopelessly non axiomatisable. Nevertheless, the two above goals have been achieved, as we shall see.

## 3. A Tentative Definition of Fuzzy Sets

We start by describing the characterisation in question with a rather informal observation. Consider an open formula $\varphi(x)$ in $\Phi$, i.e. a formula with no modal extensions, where $x$ occurs as the sole free variable. With respect to a probability structure $M$, each valuation $v$ turns $\varphi(x)$ into the analogue of a binary *random variable*. In fact, given the assumptions made, any value assignment to $x$ causes $\varphi(x)$ to have a clearly-defined probability of being true. When we consider a *set* of valuations, we obtain something similar to a binary *random field*. In our characterisation, fuzzy sets are assumed to coincide precisely with these entities.

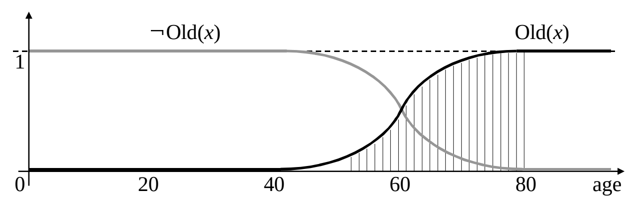
For simplicity, we will restrict our attention to monadic open formulas from this point onwards. The extension to polyadic formulas should be obvious in most cases. The machinery of generalised quantifiers enables us to express statements about fuzzy sets. For instance:

$$(x = \text{Jane} \wedge p = 0) \vee (x = \text{John} \wedge p = 0.3) \vee$$
$$(x = \text{Jill} \wedge p = 0.8) \vee (x = \text{Jack} \wedge p = 1) \ : \ \boldsymbol{P}_p \, \text{Old}(x)$$

Note that the free variables $x$ and $p$ in $\psi$ are bound in $\lambda$. Basically, the formula expresses a constraint about the fuzzy set corresponding to the (open) formula "Old($x$)". This constraint is partial, in that it relates only to a few objects which are explicitly mentioned. Another example is:

$$p = \chi_{\text{Old}}(\text{age}(x)) \ : \ \boldsymbol{P}_p \, \text{Old}(x)$$

Here, age( ) is a *measuring function* assigning an age to the objects in $D$, while $\chi_{\text{Old}}$ may be any suitable function that can be either encoded or approximated in the language of real closed fields. Note that we have assumed that all terms are rigid, so both functions do not depend on the world of reference. The function $\chi_{\text{Old}}$ acts as the membership function in a 'classical' fuzzy set, as shown in the following figure.




The function $\chi_{\text{Old}}$ defines a global constraint over the set of binary random variables, shown as vertical bars, corresponding to the valuations of $x$. Note that $\chi_{\text{Old}}$ is solely required to have [0, 1] as its domain; since it models a binary random field and not a probability distribution, such a function may consistently assign the value 1 to two or more distinct instantiations of Old($x$).

Clearly, in this scenario, fuzzy sets are given an entirely probabilistic 'deep structure'. Such a probabilistic characterisation is not new (Zimmerman 1991). The main difference, however, is that here the probabilistic trait is structurally related to a logical framework. Regarding this point, let us also observe that probability plays a role in the definition of the consequence relation only. This can also be regarded as a pragmatic advantage, as probability measures are the most constraining in the family of fuzzy measures (Klir and Yuan 1995). In other words, in the framework presented, probability leads to a stronger consequence relation.

## 4. Fuzzy Logic

In this section we will consider the most common algebraic rules for fuzzy logic from the perspective of the proposed logical framework.

**Proposition 4.1** – These three formulas are equivalent:

$p = \min(q, r)$ :
$$\forall x \, ((\boldsymbol{P}_q \, \varphi(x) \wedge \boldsymbol{P}_r \, \psi(x)) \rightarrow \boldsymbol{P}_p \, (\varphi(x) \wedge \psi(x)))$$
$p = \max(q, r)$ :
$$\forall x \, ((\boldsymbol{P}_q \, \varphi(x) \wedge \boldsymbol{P}_r \, \psi(x)) \rightarrow \boldsymbol{P}_p \, (\varphi(x) \vee \psi(x)))$$
$p = \min(1 - q + r)$ :
$$\forall x \, ((\boldsymbol{P}_q \, \varphi(x) \wedge \boldsymbol{P}_r \, \psi(x)) \rightarrow \boldsymbol{P}_p \, (\varphi(x) \rightarrow \psi(x))).$$

Furthermore, any of the above is equivalent to:

$$\forall x \, (\Box \, (\varphi(x) \rightarrow \psi(x)) \vee \Box \, (\psi \, (x) \rightarrow \varphi(x))).$$

**Proposition 4.2** – These three formulas are equivalent:

$p = \max(q + r - 1, 0)$ :
$$\forall x \, ((\boldsymbol{P}_q \, \varphi(x) \wedge \boldsymbol{P}_r \, \psi(x)) \rightarrow \boldsymbol{P}_p \, (\varphi(x) \wedge \psi(x)))$$
$p = \min(p + q, 1)$ :
$$\forall x \, ((\boldsymbol{P}_q \, \varphi(x) \wedge \boldsymbol{P}_r \, \psi(x)) \rightarrow \boldsymbol{P}_p \, (\varphi(x) \vee \psi(x)))$$
$p = \max(1 - q, r)$ :
$$\forall x \, ((\boldsymbol{P}_q \, \varphi(x) \wedge \boldsymbol{P}_r \, \psi(x)) \rightarrow \boldsymbol{P}_p \, (\varphi(x) \rightarrow \psi(x))).$$

Furthermore, any of the above is equivalent to:

$$\forall x \, (\Box \, (\varphi(x) \rightarrow \neg\psi(x)) \vee \Box \, (\neg\psi(x) \rightarrow \varphi(x))).$$

Clearly, the propositional versions of these properties also hold true. Nevertheless, the quantified sentences presented are richer in meaning due to the existing relationship between modal formulas and the sets representing the extensions of formulas in each possible world.

**Definition 4.3** – The extension in a world of an open formula $\varphi(x)$, $x$ being the only free variable in $\varphi$, is defined as:

$$\text{Ext}(w, \varphi(x)) = \{d \in D : \ (M, v[x/d], w) \models \varphi(x)\}.$$

In the light of the above definition, from a purely mathematical standpoint, every unary open formula can be taken to correspond to the analogue of a *random set*. Furthermore, all random sets of this kind share a common indexing, namely the set of possible worlds.

**Proposition 4.4 –** The formula:

$$\forall x\,(\Box\,(\varphi(x) \rightarrow \psi(x)) \vee \Box\,(\psi\,(x) \rightarrow \varphi(x)))$$

corresponds to a semantic condition of *nesting*:

$$\forall\,(w, w') \in S,\, w \neq w' \Rightarrow$$
$$(\mathrm{Ext}(w, \varphi(x) \wedge \psi(x)) \subseteq \mathrm{Ext}(w', \varphi(x) \wedge \psi(x)))\ \text{OR}$$
$$(\mathrm{Ext}(w, \varphi(x) \wedge \psi(x)) \supseteq \mathrm{Ext}(w', \varphi(x) \wedge \psi(x))).$$

**Proposition 4.5 –** The formula:

$$\forall x\,(\Box\,(\varphi(x) \rightarrow \neg\psi(x)) \vee \Box\,(\neg\psi(x) \rightarrow \varphi(x)))$$

corresponds to the semantic condition of *nesting* in Proposition 4.4 for the open formula $(\varphi(x) \wedge \neg\psi(x))$.

**Proposition 4.6 –** The formula:

$$\forall x\,\forall y\,(\Box\,(\varphi(x) \rightarrow \varphi(y)) \vee \Box\,(\varphi(y) \rightarrow \varphi(x)))$$

corresponds to the semantic condition of *nesting* in Proposition 4.4 for the open formula $\varphi(x)$.

In keeping with (Shafer 1976), the property described in Proposition 4.4 is called *joint consonance* here, whereas the property in Proposition 4.5 is called *joint dissonance*. For completeness, we also define the property in Proposition 4.6 as *consonance*.

**Proposition 4.7 –** These three formulas are equivalent:

$p = q \times r$ :
$$\forall x\,((\boldsymbol{P}_q\,\varphi(x) \wedge \boldsymbol{P}_r\,\psi(x)) \rightarrow \boldsymbol{P}_p\,(\varphi(x) \wedge \psi(x)))$$
$p = q + r - (q \times r)$:
$$\forall x\,((\boldsymbol{P}_q\,\varphi(x) \wedge \boldsymbol{P}_r\,\psi(x)) \rightarrow \boldsymbol{P}_p\,(\varphi(x) \vee \psi(x)))$$
$p = 1 - (1 - q) \times r$ :
$$\forall x\,((\boldsymbol{P}_q\,\varphi(x) \wedge \boldsymbol{P}_r\,\psi(x)) \rightarrow \boldsymbol{P}_p\,(\varphi(x) \rightarrow \psi(x))).$$

Furthermore, any of the above is equivalent to:

$$\forall x\,(\varphi(x) \perp \psi(x)).$$

The latter is an abbreviation for $(\varphi(x), \psi(x) \perp \mathit{true})$. Regarding conditional forms, we have:

**Proposition 4.8 –** The formula:

$(q = 0 \wedge p = 0) \vee (q \geq 0 \wedge p = \min(1, r / q))$ :
$$\forall x\,((\boldsymbol{P}_q\,\varphi(x) \wedge \boldsymbol{P}_r\,\psi(x)) \rightarrow \boldsymbol{P}_p\,(\varphi(x) \mid \psi(x)))$$
is equivalent to:

$$\forall x\,(\Box\,(\varphi(x) \rightarrow \psi(x)) \vee \Box\,(\psi\,(x) \rightarrow \varphi(x))).$$

**Proposition 4.9 –** The formula:

$(q = 0 \wedge p = 0) \vee (q \geq 0 \wedge p = \max(q + r - 1, 0) / q)$ :
$$\forall x\,((\boldsymbol{P}_q\,\varphi(x) \wedge \boldsymbol{P}_r\,\psi(x)) \rightarrow \boldsymbol{P}_p\,(\varphi(x) \mid \psi(x)))$$
is equivalent to:

$$\forall x\,(\Box\,(\varphi(x) \rightarrow \neg\psi(x)) \vee \Box\,(\neg\psi(x) \rightarrow \varphi(x))).$$

The ratio symbol '/' has been used for the sake of brevity.

Obviously, none of the properties presented above is a new finding. See for instance (Sales 1996) for a thorough investigation of the relationship between logic and probability. The main difference between the above algebraic rules and many-valued logics, is that the latter are taken to be fully compositional, whereas the probabilistic rules can be applied only in particular cases, i.e. when the equivalent modal conditions hold. Hájek, Godo and Esteva (1995) suggest that this difference is precisely what distinguishes the two realms; i.e. fuzzy logic is seen as a compositional theory for *degrees of truth* whereas probability is seen as a non-compositional theory of *uncertainty*. Intuitively though, the similarities between the two domains seems worth commenting on further. From a mathematical standpoint, for instance, Paris (1994) proves that the algebraic rules presented above, in a certain sense, contain all the possible definitions for many-valued logics. More precisely, he shows that any T-norm is isomorphic to either of the algebraic rules for conjunction in Propositions 4.7 and 4.2; analogously, any T-conorm is isomorphic to either of the algebraic rules for disjunction in the same propositions.

On the other hand, these algebraic rules for implication admit a further degree of freedom, depending on whether this is informally interpreted, as Lukasiewicz did (Rescher 1969), in terms of strict implication or, alternatively, as a conditional form. Note that the equivalent modal formulas are the same in both cases. Indeed, the most common choices for many-valued implications happen to fall among those described above (see Hájek and Godo 1997). Having observed that many compositional fuzzy techniques applied in practice are intentionally of a *local* nature – i.e. they are not related to a broader logical system – we come up against the suspicion that these techniques may, in fact, fall into one of the classes where compositional rules apply. In passing, Dubois and Prade (1994) are more cautious about compositionality, admitting that fuzzy sets may also be used to deal with uncertainty in a non-compositional setting, as happens in possibility theory. Nevertheless, they suggest that the formalism for degrees of truth is appropriate to fuzzy inference systems. In the next sections, we will show that these techniques may instead involve fuzzy sets of the uncertainty-based kind.

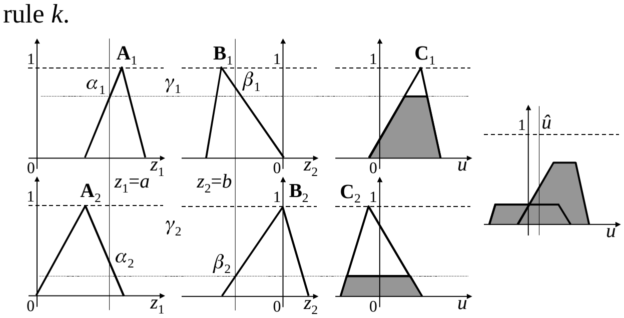
## 5. Fuzzy Inference Systems

Fuzzy inference systems are mostly, but not exclusively, used in fuzzy control systems. A comprehensive and up-to-date introduction is given in (Jang et al. 1997).

Basically, these techniques are designed to approximate a real-valued function $u = f(\boldsymbol{z})$, where $\boldsymbol{z}$ is a finite vector of real-valued parameters $z_i$. For instance, the function may describe a control signal for a system with state variables $z_i$. For simplicity, let us use a binary target function, whereby $\boldsymbol{z} = [z_1\ \ z_2]^{\mathrm{T}}$. A Mamdani-type inference system is a set of rules of the kind

**if** $z_1$ is $\mathbf{A}_k$ **and** $z_2$ is $\mathbf{B}_k$ **then** $u$ is $\mathbf{C}_k$

where $\mathbf{A}_k$, $\mathbf{B}_k$ and $\mathbf{C}_k$ are fuzzy subsets of the real axis in rule $k$.



The inference method for two such rules is visually described in the above figure. The vertical bars correspond to the two values $a$ and $b$ for the parameters $z_1$ and $z_2$ respectively. These values are intersected with the fuzzy sets; the intersection values $\alpha_k$ and $\beta_k$ are then combined through a T-norm (e.g. min) to obtain the values $\gamma_k$. The latter are used as thresholds to 'cut' the fuzzy sets $C_k$; 'cut' sets originating from different rules are combined through a T-conorm (e.g. max) to yield a final fuzzy set. The estimated value $\hat{u}$ is finally obtained through a *defuzzification* method; typically through an average-like operation.

In a Sugeno-type inference system, fuzzy rules are of a slightly different kind:

**if** $z_1$ is $\mathbf{A}_k$ **and** $z_2$ is $\mathbf{B}_k$ **then** $u$ is $f_k(z_1, z_2)$.

Each $f_k$ is intended to be a local approximation to the function $f$. The values $\gamma_k$ are computed in the same way as with Mamdani-type rules. The main difference resides in how the estimated value $\hat{u}$ is computed, typically as the 'average' of the contributions given by the $f_k$:

$$\hat{u} = \gamma_1 f_1(a, b) + \gamma_2 f_2(a, b) + \ldots + \gamma_n f_n(a, b).$$

It can be proven that both inference systems, under certain conditions, are universal approximators; i.e. the value $\hat{u}$ can be made arbitrarily close to any continuous target function (Buckley, 1995). In the field of practical applications, fuzzy approximation techniques are particularly appreciated when the target function is either unknown or unobservable, as they apparently relate more directly to human knowledge.

The logical encoding of the above techniques can be made clearer by establishing an informal relation with a possibly more familiar probabilistic model. Assuming that the target function is unobservable, we may model the variable $u$ and the parameters $z_i$ as random variables on a suitable sample space $\Omega$. The conditional density $P(u \mid \boldsymbol{z})$ would then describe the probability for $u$ to be the 'true' value of the function, given the parameter vector $\boldsymbol{z}$. In the logical encoding we assume a bijective correspondence in meaning between the objects $d \in D$ and the sample points $\omega \in \Omega$. For instance, if the target function describes a control strategy, each object/point represents a possible state of the physical system being controlled.

The encoding of the fuzzy sets which occur in the rules is exemplified as follows:

$$p = \chi_{A_k}(z_1(x)) \;:\; \boldsymbol{P}_p\, A_k(x)$$
$$p = \chi_{B_k}(z_2(x)) \;:\; \boldsymbol{P}_p\, B_k(x)$$
$$p = \chi_{C_k}(u(x)) \;:\; \boldsymbol{P}_p\, C_k(x).$$

Here, $z_1$, $z_2$ and $u$ are encoded as *measuring functions* assigning a real value to the objects in $D$. In passing, measuring functions are used in (Bacchus 1990) to represent the analogue of random variables. The analogy does not hold here since there is no 'randomness' on the variable $x$, i.e. probability distributes on $W$ rather than on $D$. Mamdani-type rules are encoded as strict implications:

$$\forall x\, \Box\, ((A_k(x) \wedge B_k(x)) \rightarrow C_k(x)).$$

An entire rule base is encoded as the overall disjunction of the above rules. Any specific value assignment to the parameters $z_1$ and $z_2$ can be expressed through a restriction:

$$(z_1(x) = a) \wedge (z_2(x) = b)$$

which circumscribes the relevant objects in $D$. There is a crucial analogy here between the latter circumscription and the *conditional* update of a probability space. In a probability space, the acquisition of new facts causes the elimination of the sample points in $\Omega$ which have become irrelevant. This forces the updating of $P(\,)$ into the conditional form $P(\,\mid z_1 = a,\, z_2 = b)$. Similarly, the above circumscription corresponds to the elimination of the irrelevant objects from $D$. Nevertheless, the measure $\mu$ on possible worlds and hence the probability operators are unaffected since no worlds are ruled out.

The intersection values $\alpha_k$ and $\beta_k$ are algebraically computed from the conjunctions:

$$p = \chi_{A_k}(z_1(x)) \wedge (z_1(x) = a) \;:\; \boldsymbol{P}_p\, A_k(x)$$
$$p = \chi_{B_k}(z_2(x)) \wedge (z_2(x) = b) \;:\; \boldsymbol{P}_p\, B_k(x).$$

The min operator is a sound choice in this case, as it represents the 'erosion' of irrelevant objects from the original fuzzy sets. A quite different matter is the combination of $\alpha_k$ and $\beta_k$ to obtain $\gamma_k$. Observe that the min operator – i.e. the most popular choice – is only applicable when the modal condition in Proposition 4.1 holds between $A_k(x)$ and $B_k(x)$. This also means that the two fuzzy sets would have to be jointly consonant. Let us provisionally assume that this is true of all pairs $A_k$ and $B_k$; the implication of this will be discussed shortly. It can be proved that:

$$p \;:\; \forall x\, (\Box\, (\varphi(x) \rightarrow \psi(x)) \rightarrow (\Box_p\, \varphi(x) \rightarrow \Box_p\, \psi(x)))$$
$$p = \max(q, r) :$$
$$\forall x\, ((\Box_q\, \varphi(x) \wedge \Box_r\, \psi(x)) \rightarrow \Box_p\, (\varphi(x) \vee \psi(x)))$$

are valid formulas. The first makes it possible to derive the formula describing the 'cut' fuzzy sets:

$$(p = \max(\chi_{C_k}(u(x)), \gamma_k)) \wedge (z_1(x) = a) \wedge (z_2(x) = b) :$$
$$\Box_p\, C_k(x)$$

i.e. a lower bound on probabilities. The second valid formula makes it possible to combine the 'cut' fuzzy sets arising from different rules through the max operator, thus obtaining the final fuzzy set; i.e. a cumulative lower bound.

In this light, the identification of $\hat{u}$ through an 'averaging' operation is vaguely reminiscent of the calculation of an expected value. However, the fuzzy set in question is not a probability distribution, nor even a lower one, so this choice seems difficult to justify.

Sugeno-type rules are encoded in the following way:

$$V_k(x) \equiv \mathrm{Approx}f(u(x),\; f_k(z_1(x), z_2(x))) \;:$$
$$\forall x\, \Box\, ((A_k(x) \wedge B_k(x)) \rightarrow V_k(x))$$

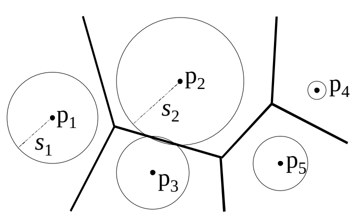
where $V_k$ is an auxiliary predicate and Approx$f$ describes a viable approximation to $f$. When applied to input values, each of these rules yields a point-wise estimate of $\hat{u}$, thus the final 'expected value' can be obtained directly. One critical point in Sugeno's technique, however, is precisely this 'expected value'. In a certain sense, $\{\gamma_k\}$ appears to be quite similar to a conditional, discrete probability distribution over the possible values of $\hat{u}$. If this is the case, the expectation operator would be appropriate. This point leads us to discussing a fundamental aspect of informal interpretation.

## 6. A Short Epistemological Interlude

Intuitively, the techniques in the previous section require proper coverage of the domain of the target function through fuzzy sets such as $A_k$ and $B_k$. In passing, the notion of a fuzzy partition is usually defined in the literature by a set of natural axioms accounting for what in this respect is proper coverage. From another point of view, Definition 4.3 says that fuzzy sets such as $A_k$ and $B_k$ have a 'deep structure' in terms of random sets. Technical details apart, the main question here regards the informal rationale for uncertainty in a structure of related predicate extensions.

A few helpful ideas for providing an answer have been formulated by P. Gärdenfors. In (1997, 1992), with the support of some psychological experiments, he theorises an intermediate level of information representation between the 'symbolic' level – namely the level of predicates – and what he calls the 'sub-conceptual' level, i.e. the kind of associative, intrinsic representation proper to neural networks. This kind of 'missing link', called the 'conceptual' level, is designed to explain how concept formation may take place. A *conceptual space* is a number of *quality dimensions*. For instance, dimensions of these kinds may be closely related to human sensory receptors, such as spatial dimensions, temperature, colours, etc. The term 'dimension' is to be understood in its proper mathematical sense; conceptual spaces are taken as being endowed with a geometric or topological structure. In this abstraction, properties are regions, possibly convex; concepts – corresponding to predicates – are either a property or a set of properties defined on different dimensions.

Gärdenfors also adopts a theory of *prototypes*, i.e. highly representative points acting as attractors in a conceptual space, around which concepts are potentially formed. As one formal pattern for concept formation, he considers a variation of Voronoï tesselations called *power diagrams* (Aurenhammer 1991). An example of a power diagram is given in the figure:




A Voronoï tessellation is a topological abstraction based on the notion of distance, which applies to many natural phenomena, such as crystal growth. Given a metric space and a set of characteristic points – e.g. the prototypes in this case – the Voronoï cell around each characteristic point is defined as the region closer to it than to any other characteristic point. With power cells, every characteristic point is also endowed with its own *strength*, as represented by the radius of a hypersphere centred in it. Formally, the boundary equation between two contiguous cells is:

$$(x - \mathrm{p}_i)^{\mathrm{T}}(x - \mathrm{p}_i) - s_i^2 = (x - \mathrm{p}_j)^{\mathrm{T}}(x - \mathrm{p}_j) - s_j^2$$

where $s_i$ is the strength of point $\mathrm{p}_i$. Voronoï tesselations correspond to the case when all strengths $s_i$ are zero. In all other cases, distances are measured from the borders of the power hypersphere rather than from the points themselves. In passing, both Voronoï and power cells are always convex and divided by straight lines. Of course, shapes other than a sphere might be considered for further generality.

In such a scenario, we propose that uncertainty could be construed as yet another dimension relating to the relative strengths of prototypes. In other words, a symbolic system in a conceptual space might conveniently be modelled as a family of 'possible tessellations' where the prototypes are fixed and their relative strengths vary. In turn, the addition of (subjective) probability makes the symbolic system to correspond to a random tessellation where each concept is represented by a random set.

The proposal can be further clarified by relating it to our main topic. Let us consider a Sugeno-type inference system with a set of $k$ prototypical points in the domain of the target function $f$ for which suitable approximations $f_k$ are known. For instance, $f$ could again describe the control signal for a physical system, with each $f_k$ being a local control strategy whose effectiveness has been experimentally assessed. In the logical encoding, the 'concepts' – i.e. the extension of predicates $V_k$ – are simply the regions where the corresponding approximation applies. In this respect, the adoption of just one such tessellation is very committing from an informal standpoint and leads to technical difficulties. On the other hand, fuzzy inference systems may be taken as proving that an effective approximation technique with a low computational cost can be achieved by adopting random tessellations.

In this light, uncertainty may be held to add flexibility to a conceptual space in that it improves the pragmatic applicability of a symbolic system defined in it. It is crucial to observe, however, that uncertainty comes into play when a *purpose* is attached to a conceptual tessellation, or, to put the matter differently, when the symbols associated to the concepts are inserted into the dynamics of a reasoning process, such as that of identifying a purposeful approximation to an unknown function. Informally, this is quite different from uncertainty about actual facts, as for instance in the random model of a noisy sensor. In the latter case, randomness stems from the sum of the erratic effects in the sensing system, whereas randomness in a conceptual space is taken to be due to the purposeful application of a complex of symbols in a reasoning process.

Inter alia, this also explains why the actual modelling of fuzzy sets is commonly reported to be so strongly dependent on context. The shape of a fuzzy set such as Old($x$) depends on what is taken to be entailed from accepting that a an object is 'Old' in a reasoning context. Likewise, fully-shaped fuzzy sets can hardly make sense outside the reasoning system in which they are originated. The main advantage of a logical framework for this analysis is to make this systemic nature of fuzzy sets much more evident.

## 7. Fuzzy Partitions, Topological Reasoning

We can now turn to the logical encoding of fuzzy partitions by discussing the use of random tessellations for the underlying 'deep structure'. Let $\mathrm{V}_k$ denote a finite set of $n$ predicates describing a random tessellation. Two main axiom schema apply:

a) $\forall x\ \Box\, (\mathrm{V}_1(x) \lor \ldots \lor \mathrm{V}_n(x))$
b) $\forall x\ \Box\, (\mathrm{V}_i(x) \to \neg \mathrm{V}_j(x)), \quad \forall i,j \in [1..n],\ i \neq j$ .

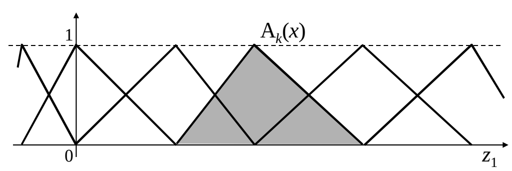
These axioms mean that each possible world in $S$ contains the complete description of a tessellation of $D$.

In the light of the previous section, we may further assume that a conceptual space is encoded by a suitable set of measuring functions corresponding to each dimension. The idea of concepts as convex regions is straightforwardly extensible to the random dimension by adopting random power tessellations. Two more properties derive from this:

c) $\exists x\ \Box\, \mathrm{V}_i(x)$
d) $\forall x \forall y\ (\Box\, (\mathrm{V}_i(x) \to \mathrm{V}_i(y)) \lor \Box\, (\mathrm{V}_i(y) \to \mathrm{V}_i(x)))$

which hold for any $i \in [1..n]$. Property c) states that each 'concept' $\mathrm{V}_k$ has at least one prototype, which also means that fuzzy sets $\mathrm{V}_k$ are normal – i.e. their membership functions reaches unity for some objects. Property d) states that fuzzy sets $\mathrm{V}_k$ are *consonant*.

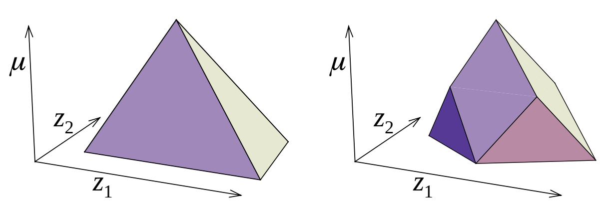
For obvious computational reasons, fuzzy partitions are often taken in practice to be decomposable, i.e. to be expressible as the combination of fuzzy partitions on each dimension. For instance, a monodimensional fuzzy partition may have the following shape:



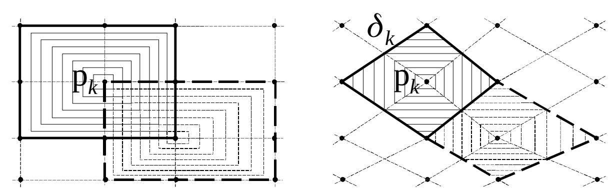

As already stated, the min T-norm can be consistently applied to combinations of fuzzy sets which are jointly consonant. At first sight, this seems to support an informal 'psychological' preference for convex tessellating elements. In the case of the example in Section 5, decomposability plus joint consonance would entail that

$$p = \min(q, r) : \forall x\ ((\boldsymbol{P}_q\ \mathrm{A}_k(x) \land \boldsymbol{P}_r\ \mathrm{B}_k(x)) \to \boldsymbol{P}_p\ \mathrm{V}_k(x)).$$

as we provisionally assumed in Section 5.




However, this is incompatible with a random tessellation. In fact, when the two dimensions are orthogonal and $\mathrm{A}_k$ and $\mathrm{B}_k$ are triangularly shaped, the tessellating element for the three-dimensional space – $z_1$ and $z_2$ plus the 'random' axis – is a kind of pyramid, as in the left part of the above figure. Clearly, such a pyramid is not a tessellating element for the three-dimensional space in question. Accordingly, it is easily proved that no random power tessellation exists where rectangles are the only possible shapes. Hence complete coverage with nested rectangles of varying dimensions can only be achieved by allowing the cells to overlap in some worlds, thus violating axiom b) above. This situation is represented in the left hand part of the following figure:




Consequently, the set of values $\gamma_k$ defined in Section 5 is not a discrete probability distribution since, due to the overlapping of rectangles, the sum will generally exceed one. Nevertheless, the use of max is always sound in Mamdani-type rules if lower probability limits are intended; the price to pay, however, is that it is not possible to infer point-valued constraints. A similar line of reasoning applies to the product T-norm as well – i.e. the second most popular choice – with the further problem that not even convexity is preserved. In turn this makes it difficult to justify informally why the prototypical elements, i.e. the vertices of triangles, are still there.

Interestingly, the same reasoning can be used to find a random power tessellation – based on a generalised distance – admitting a simple algebraic rule for conjunction. Without going into details, the random tessellation in question is represented on the right hand side of the figure above. In this case, the varying rectangles are allowed to mutate into more complex shapes to avoid overlapping. The tessellating element for the three-dimensional space is given above, on the right hand side of the pyramid. In this case, the monodimensional fuzzy sets are construed as the projections of the random tessellation onto each axis. The corresponding compositional rule is:

$$((\delta_k(x) = 1) \land (p = \min(q, r)) \lor ((\delta_k(x) = 0) \land (p = 0)\ :$$
$$\forall x\ ((\boldsymbol{P}_q\ \mathrm{A}_k(x) \land \boldsymbol{P}_r\ \mathrm{B}_k(x)) \to \boldsymbol{P}_p\ \mathrm{V}_k(x))$$

where $\delta_k$ is the characteristic function of the bounding 'diamond' for $\mathrm{V}_k$. Note that in general this is not a T-norm, unless the diamond is a square. However, the net advantage is that the set of values $\gamma_k$ computed with the above rule is now a conditional, discrete distribution over the possible estimates for $\hat{u}$.

As we have seen, analysis of the 'deep structure' of a fuzzy partition throws a different light on the meaning of algebraic rules. Note also that these aspects are totally invisible when the very same algebraic rules are studied in isolation. Instead, the adoption of a structured characterisation for fuzzy sets brings to light an underlying 'deep' level of topological reasoning that also emerges with Gärdenfors' conceptual spaces.

## 8. Conclusions and Future Work

One aspect which has not been discussed in this paper is how the uncertainty model for fuzzy sets presented ties up with other uncertainty models regarding factual phenomena, e.g. the random model of a noisy sensor. Models of the latter kind, from the standpoint of the logical framework, involve randomness over the object domain $D$ and hence require the introduction of a second probability measure on $D$. Thus, the fuzzy inference systems discussed in Section 5 could be extended to embrace the case where the parameter values are described by a probability density.

The extension of the formalism to include a second probability measure has already been contemplated in (Halpern 1990) and formally studied in a number of subsequent works. In these studies, however, the informal objective is somewhat different from what we are proposing here. In keeping with a long-standing tradition dating back to Carnap, the two measures are held to represent two different kinds of probability, namely statistical probability – i.e. on $D$ – and degrees of belief – i.e. on $W$. In our line of thinking, a slightly different direction seems appropriate, namely conceiving a unique probability space where the two forms of uncertainty, one relating to factual phenomena and the other intrinsic to the purposive use of a symbolic system, are brought together. It might be assumed that the unique probability space is decomposable into the two independent measures mentioned above. However, in our understanding, the assumption that the learning of new facts does not in any way alter a systemic fuzzy set model cannot be taken for granted. Hopefully, the assessment of the interactions may instead help to achieve a formal account for more complex, 'gestalt'-like phenomena.

As we have seen, the logical framework presented may provide a valuable formal tool for investigating fuzzy models in terms of their 'deep structures'. Maybe this will provide a safer bridge over the gap between the realm of fuzziness and that of probability.

**Acknowledgements**. Thanks are due to Didier Dubois, Université Paul Sabatier, Toulouse, and Rosella Gennari, University of Amsterdam, for their precious advice and suggestions.

## References

Abadi, M., Halpern, J. Y., 1994, Decidability and expressiveness for first-order logics of probability, *Information and Computation*, 112 (1), 1-36.

Aurenhammer, F., 1991, Voronoi diagrams: A survey of a fundamental geometric data structure, *ACM Computing Surveys*, 23, 345-405.

Bacchus, F., 1990, *Representing and Reasoning with Probabilistic Knowledge*, Cambridge, MA: The MIT Press.

Buckley, J. J., 1995, System Stability and the Fuzzy Controller, in Nguyen, H. T. et al. (eds.), *Theoretical Aspects of Fuzzy Control*, New York: John Wiley & Sons.

Dubois, D., Prade, H., 1994, Can We Enforce Full Compositionality in Uncertainty Calculi? In *Proceedings of the 12th National Conference on Artificial Intelligence*, Vol. 1, 149-154, Menlo Park, CA: AAAI Press.

Gaifman, H., Snir, M., 1982, Probabilities over rich languages, testing and randomness, *The Journal of Symbolic Logic*, 47 (3), 495-548.

Gärdenfors, P., 1992, A geometric model of concept formation, in S. Ohsuga et al. (eds.), *Information Modelling and Knowledge Bases III*, 1-16, Amsterdam: IOS Press.

Gärdenfors, P., 1997, Symbolic, conceptual and subconceptual representations, in Cantoni, V., et al. (eds.), *Human and Machine Perception: Information Fusion*, 255-270, New York: Plenum Press.

Gerla, G., 1994, Inferences in probability logic, *Artificial Intelligence*, 70, 33-52.

Hájek, P., Godo, L., 1997, Deductive systems of fuzzy logic – a tutorial, Technical Report no. 707, Inst. of Computer Science, Academy of Sciences of the Czech Republic.

Hájek, P., Godo, L., Esteva, F., 1995, Fuzzy Logic and Probability, *Proceedings of the 11th Annual Conference on Uncertainty in Artificial Intelligence*, San Francisco, CA: Morgan Kaufmann Publishers.

Halpern, J. Y., 1990, An analysis of first-order logics of probability, *Artificial Intelligence* 46, 1990, 311-350.

Jang, J. S. R., Sun, C. T., Mizutani, E., 1997, *Neuro-Fuzzy and Soft Computing*, Upper Saddle River, NJ: Prentice Hall.

Klir, G. J., Yuan, B., 1995, *Fuzzy Sets and Fuzzy Logic*, Upper Saddle River, NJ: Prentice Hall.

Paris, J. B., 1994, *The Uncertain Reasoner's Companion*, Cambridge, UK: Cambridge University Press.

Rescher, N., 1969, *Many-Valued Logics*, New York: Mc Graw-Hill.

Ruspini, E. H., 1991, On the Semantics of Fuzzy Logic, *International Journal of Approximate Reasoning*, 5, 45-88.

Sales, T., 1996, Logic as General Rationality: A Survey, in Gabbay, D., Guenthner, F. (eds.), *Handbook of Philosophical Logic, 2nd ed.*, Forthcoming.

Shafer, G., 1976, *A mathematical theory of evidence*, Princeton, NJ : Princeton University Press.

Voorbraak, F. 1993. As Far as I Know, Ph.D. diss., Dept. of Philosophy, Utrecht Univ.

Wang, P., Tan, S., 1997, *Soft Computing and Fuzzy Logic*, Soft Computing, 1, 35-41.

Zadeh, L.A., 1965, Fuzzy sets. *Information and Control*, 8, 338-353.

Zadeh, L.A., 1978, Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1, 3-28.

Zimmerman, H. J., 1991, *Fuzzy Set Theory and Its Applications, 2nd ed.*, Boston, MA: Kluwer Academic Publishers.