

Chapter 2

VAGUE CONCEPTS AND FUZZY SETS

Vague or fuzzy concepts are fundamental to natural language, playing a central role in communications between individuals within a shared linguistic context. In fact Russell [90] even goes so far as to claim that all natural language concepts are vague. Yet often vague concepts are either viewed as problematic because of their susceptibility to Sorities paradoxes or at least as somehow ‘second rate’ when compared with the more precise concepts of the physical sciences, mathematics and formal logic. This view, however, does not properly take account of the fact that vague concepts seem to be an effective means of communicating information and meaning. Sometimes more effective, in fact, than precise alternatives. Somehow, knowing that ‘the robber was *tall*’ is more useful to the police patrolling the streets, searching for suspects, than the more precise knowledge that ‘the robber was *exactly 1.8 metres* in height’. But what is the nature of the information conveyed by fuzzy statements such as ‘the robber was *tall*’ and what makes it so useful? It is an attempt to answer this and other related questions that will be the central theme of this volume. Throughout, we shall unashamedly adopt an Artificial Intelligence perspective on vague concepts and not even attempt to resolve longstanding philosophical problems such as Sorities paradoxes. Instead, we will focus on developing an understanding of how an intelligent agent can use vague concepts to convey information and meaning as part of a general strategy for practical reasoning and decision making. Such an agent could be an artificial intelligence program or a human, but the implicit assumption is that their use of vague concepts is governed by some underlying internally consistent strategy or algorithm. For simplicity this agent will be referred to using the pronoun You. This convention is borrowed from Smets work on the Transferable Belief Model (see for example [97]) although the focus of this work is quite different. We shall immediately attempt to reduce the enormity of our task by restricting the type of vague

concept to be considered. For the purposes of this volume we shall restrict our attention to concepts as identified by words such as adjectives or nouns that can be used to describe a object or instance. For such an expression θ it should be meaningful to state that ‘ x is θ ’ or that ‘ x is a θ ’¹. Given a universe of discourse Ω containing a set of objects or instances to be described, it is assumed that all relevant expression can be generated recursively from a finite set of basic labels LA . Operators for combining expressions are restricted to the standard logical connectives of negation (\neg), conjunction (\wedge), disjunction (\vee) and implication (\rightarrow). Hence, the set of label expressions identifying vague concepts can be formally defined as follows:

DEFINITION 1 *Label Expressions*

The set of label expressions of LA , LE , is defined recursively as follows:

(i) $L \in LE, \forall L \in LA$

(ii) If $\theta, \varphi \in LE$ then $\neg\theta, \theta \wedge \varphi, \theta \vee \varphi, \theta \rightarrow \varphi \in LE$

For example, Ω could be the set of suspects for a robbery and LA might correspond to a set of basic labels used by police for identifying individuals, such as $LA = \{tall, medium, short, medium\ build, heavy\ build, stocky, thin, \dots, blue\ eyes, brown\ eyes \dots\}$. In this case possible expressions in LE include $medium \wedge \neg tall \wedge brown\ eyes$ (‘medium but not tall with brown eyes’) and $short \wedge (medium\ build \vee heavy\ build)$ (‘short with medium or heavy build’).

Since it was first proposed by Zadeh in 1965 [110] the treatment of vague concepts in artificial intelligence has been dominated by fuzzy set theory. In this volume, we will argue that aspects of this theory are difficult to justify, and propose an alternative perspective on vague concepts. This in turn will lead us to develop a new mathematical framework for modelling and reasoning with imprecise concepts. We begin, however, in this first chapter by reviewing current theories of vague concepts based on fuzzy set theory. This review will take a semantic, rather than purely axiomatic, perspective and investigate a number of proposed operational interpretations of fuzzy sets, taking into account their consistency with the mathematical calculus of fuzzy theory.

2.1 Fuzzy Set Theory

The theory of fuzzy sets, based on a truth-functional calculus proposed by Zadeh [110], is centred around the extension of classical set theoretic operations such as union and intersection to the non-binary case. Fuzzy sets are generalisations of classical (crisp) sets that allow elements to have partial membership. Every crisp set A is characterised by its membership function $\chi_A : \Omega \rightarrow \{0, 1\}$ where $\chi_A(x) = 1$ if and only if $x \in A$ and where $\chi_A(x) = 0$ otherwise. For

fuzzy sets this definition is extended so that $\chi_A : \Omega \rightarrow [0, 1]$ allowing x to have partial membership $\chi_A(x)$ in A .

Fuzzy sets can be applied directly to model vague concepts through the notion of extension. The extension of a crisp (non-fuzzy) concept θ is taken to be those objects in the universe Ω which satisfy θ i.e $\{x \in \Omega : 'x \text{ is } \theta' \text{ is true}\}$. In the case of vague concepts it is simply assumed that some elements have only partial membership in the extension. In other words, the extension of a vague concept is taken to be a fuzzy set. Now in order to avoid any cumbersome notation we shall also use θ to denote the extension of an expression $\theta \in LE$. Hence, according to fuzzy set theory [110] the extension of a vague concept θ is defined by a fuzzy set membership function $\chi_\theta : \Omega \rightarrow [0, 1]$. Now given this possible framework You are immediately faced with a difficult computational problem. Even for a finite basic label set LA there are infinitely many expressions in LE generated by the recursive definition 1. You cannot hope to explicitly define a membership function for any but a small subset of these expressions. Fuzzy set theory attempts to overcome this problem by providing a mechanism according to which the value for $\chi_\theta(x)$ can be determined uniquely from the values $\chi_L(x) : L \in LE$. This is achieved by defining a mapping function for each of the standard logical connectives; $f_\wedge : [0, 1]^2 \rightarrow [0, 1]$, $f_\vee : [0, 1]^2 \rightarrow [0, 1]$, $f_\rightarrow : [0, 1]^2 \rightarrow [0, 1]$ and $f_\neg : [0, 1] \rightarrow [0, 1]$. The value of $\chi_\theta(x)$ for any expression θ and value $x \in \Omega$ can then be determined from $\chi_L(x) : L \in LA$ according to the following recursive rules:

$$\forall \theta, \varphi \in LE \chi_{\theta \wedge \varphi}(x) = f_\wedge(\chi_\theta(x), \chi_\varphi(x))$$

$$\forall \theta, \varphi \in LE \chi_{\theta \vee \varphi}(x) = f_\vee(\chi_\theta(x), \chi_\varphi(x))$$

$$\forall \theta, \varphi \in LE \chi_{\theta \rightarrow \varphi}(x) = f_\rightarrow(\chi_\theta(x), \chi_\varphi(x))$$

$$\forall \theta \in LE \chi_{\neg \theta}(x) = f_\neg(\chi_\theta(x))$$

This assumption is referred to as truth-functionality ² due to the fact that it extends the recursive mechanism for determining the truth-values of compound sentences from propositional variables in propositional logic. In fact, a fundamental assumption of fuzzy set theory is that the above functions coincide with the classical logic operators in the limit case when $\chi_\theta(x), \chi_\varphi(x) \in \{0, 1\}$. Beyond this constraint it is somewhat unclear as to what should be the precise definition of these combination functions. However, there is a wide consensus that f_\wedge, f_\vee and f_\neg [54] should satisfy the following sets of axioms:

Conjunction

$$\mathbf{C1} \quad \forall a \in [0, 1] f_\wedge(a, 1) = a$$

$$\mathbf{C2} \quad \forall a, b, c \in [0, 1] \text{ if } b \leq c \text{ then } f_\wedge(a, b) \leq f_\wedge(a, c)$$

$$\mathbf{C3} \quad \forall a, b \in [0, 1] \quad f_{\wedge}(a, b) = f_{\wedge}(b, a)$$

$$\mathbf{C4} \quad \forall a, b, c \in [0, 1] \quad f_{\wedge}(f_{\wedge}(a, b), c) = f_{\wedge}(a, f_{\wedge}(b, c))$$

Disjunction

$$\mathbf{D1} \quad \forall a \in [0, 1] \quad f_{\vee}(a, 0) = a$$

$$\mathbf{D2} \quad \forall a, b, c \in [0, 1] \quad \text{if } b \leq c \text{ then } f_{\vee}(a, b) \leq f_{\vee}(a, c)$$

$$\mathbf{D3} \quad \forall a, b \in [0, 1] \quad f_{\vee}(a, b) = f_{\vee}(b, a)$$

$$\mathbf{D4} \quad \forall a, b, c \in [0, 1] \quad f_{\vee}(f_{\vee}(a, b), c) = f_{\vee}(a, f_{\vee}(b, c))$$

Negation

$$\mathbf{N1} \quad f_{\neg}(1) = 0 \text{ and } f_{\neg}(0) = 1$$

$$\mathbf{N2} \quad f_{\neg} \text{ is a continuous function on } [0, 1]$$

$$\mathbf{N3} \quad f_{\neg} \text{ is a decreasing function on } [0, 1]$$

$$\mathbf{N4} \quad \forall a \in [0, 1] \quad f_{\neg}(f_{\neg}(a)) = a$$

Axioms C1-C4 mean that f_{\wedge} is a triangular norm or (t-norm) as defined by [94] in the context of probabilistic metric spaces. Similarly according to D1-D4 f_{\vee} is a triangular conorm (t-conorm). An infinite family of functions satisfy the t-norm and t-conorm axioms including $f_{\wedge} = \min$ and $f_{\vee} = \max$ proposed by Zadeh [110]. Other possibilities are, for conjunction, $f_{\wedge}(a, b) = a \times b$ and $f_{\wedge}(a, b) = \max(0, a + b - 1)$ and, for disjunction, $f_{\vee}(a, b) = a + b - a \times b$ and $\min(1, a + b)$. Indeed it can be shown [54] that f_{\wedge} and f_{\vee} are bounded as follows:

$$\forall a, b \in [0, 1] \quad \underline{f}_{\wedge}(a, b) \leq f_{\wedge}(a, b) \leq \min(a, b)$$

$$\forall a, b \in [0, 1] \quad \max(a, b) \leq f_{\vee}(a, b) \leq \overline{f}_{\vee}(a, b)$$

where \underline{f}_{\wedge} is the drastic t-norm defined by:

$$\forall a, b \in [0, 1] \quad \underline{f}_{\wedge}(a, b) = \begin{cases} a & : b = 1 \\ b & : a = 1 \\ 0 & : \text{otherwise} \end{cases}$$

and \overline{f}_{\vee} is the drastic t-conorm defined by:

$$\forall a, b \in [0, 1] \quad \overline{f}_{\vee}(a, b) = \begin{cases} a & : b = 0 \\ b & : a = 0 \\ 1 & : \text{otherwise} \end{cases}$$

Interestingly, adding the additional idempotence axioms restricts t-norms to min and t-conorms to max:

Idempotence Axioms

C5 $\forall a \in [0, 1] f_{\wedge}(a, a) = a$

D5 $\forall a \in [0, 1] f_{\vee}(a, a) = a$

THEOREM 2 f_{\wedge} satisfies C1-C5 if and only if $f_{\wedge} = \min$

Proof

(\Leftarrow) Trivially min satisfies C1-C5

(\Rightarrow) Assume f_{\wedge} satisfies C1-C5

For $a, b \in [0, 1]$ suppose $a \leq b$ then

$$a = f_{\wedge}(a, a) \leq f_{\wedge}(a, b) \leq f_{\wedge}(a, 1) = a$$

by axioms C1, C2 and C5 and therefore $f_{\wedge}(a, b) = a = \min(a, b)$

Alternatively, for $a, b \in [0, 1]$ suppose $b \leq a$ then

$$b = f_{\wedge}(b, b) \leq f_{\wedge}(a, b) \leq f_{\wedge}(1, b) = b$$

by axioms C1, C2, C3 and C5 and therefore $f_{\wedge}(a, b) = b = \min(a, b) \square$

THEOREM 3 f_{\vee} satisfies D1-D5 if and only if $f_{\vee} = \max$

The most common negation function f_{\neg} proposed is $f_{\neg}(a) = 1 - a$ although there are again infinitely many possibilities including, for example, the family of parameterised functions defined by:

$$f_{\neg}^{\lambda}(a) = \frac{1 - a}{1 + \lambda a}$$

Somewhat surprisingly, however, all negation functions essentially turn out to be rescalings of $f_{\neg}(a) = 1 - a$ as can be seen from the following theorem due to Trillas [101].

THEOREM 4 If f_{\neg} satisfies N1-N4 then $\langle [0, 1], f_{\neg}, < \rangle$ is isomorphic to $\langle [0, 1], 1 - x, < \rangle$

Proof

Since $f_{\neg}(1) = 0$ and $f_{\neg}(0) = 1$ then by continuity (N2) it follows that there is some value k such that $f_{\neg}(k) = k$ (see figure 2.1). Now define

$$g(x) = \begin{cases} \frac{x}{2k} & : x \leq k \\ 1 - \frac{f_{\neg}(x)}{2k} & : x > k \end{cases}$$

Then $g(0) = 0$, $g(1) = 1$ and $g(k) = 0.5$. Also, it is easy to check that g is strictly increasing and continuous and hence onto.

Finally, for $x \leq k$ $f_{\neg}(x) \geq f_{\neg}(k) = k$ (by N3) and therefore

$$g(f_{\neg}(x)) = 1 - \frac{f_{\neg}(f_{\neg}(x))}{2k} = 1 - \frac{x}{2k} = 1 - g(x) \text{ by N4}$$

Similarly, for $x > k$ then $f_{\neg}(x) \leq f_{\neg}(k) = k$ and therefore

$$g(f_{\neg}(x)) = \frac{f_{\neg}(x)}{2k} = 1 - g(x) \text{ as required. } \square$$

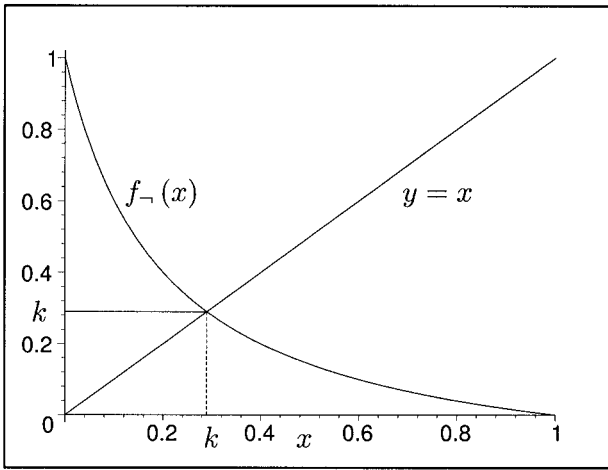


Figure 2.1: Plot of a possible f_{\neg} function and its associated k value

In view of this rather strong result we shall now assume that $f_{\neg}(a) = 1 - a$ and move on to consider possible relationships between t-norms and t-conorms. Most of the constraints relating t-norms and t-conorms come from the imposition of classical logic equivalences on vague concepts. Typical of these is the duality relationship that emerges from the assumption that vague concepts satisfy de Morgan's Law i.e. that $\theta \vee \varphi \equiv \neg\theta \wedge \neg\varphi$. In the context of truth-functional fuzzy set theory this means that:

$$\forall a, b \in [0, 1] f_{\vee}(a, b) = 1 - f_{\wedge}(1 - a, 1 - b)$$

Accordingly the table in figure 2.2 shows a number of well known t-norms with their associated t-conorm duals:

Another obvious choice of logical equivalence that we might wish vague concepts to preserve is $\theta \wedge \theta \equiv \theta$ (similarly $\theta \vee \theta \equiv \theta$). In terms of t-norms (t-conorms) this leads to the idempotence axiom C5 (D5) which as we have seen from theorem 2 (theorem 3) restricts us to min (max).

t-norm $f_{\wedge}(a, b)$	t-conorm dual $f_{\vee}(a, b)$
$\min(a, b)$	$\max(a, b)$
$a \times b$	$a + b - a \times b$
$\max(0, a + b - 1)$	$\min(1, a + b)$
$f_{\wedge}(a, b)$	$f_{\vee}(a, b)$

Figure 2.2: t-norms with associated dual t-conorms

In addition, to constraints based on classical logical equivalences it might also be desirable for fuzzy memberships to be additive in the sense that

$$\forall x \in \Omega \forall \theta, \varphi \in LE \chi_{\theta \vee \varphi}(x) = \chi_{\theta}(x) + \chi_{\varphi}(x) - \chi_{\theta \wedge \varphi}(x)$$

This generates the following equation relating t-norms and t-conorms:

$$\forall a, b \in [0, 1] f_{\vee}(a, b) = a + b - f_{\wedge}(a, b)$$

Making the additional assumption that f_{\vee} is the dual of f_{\wedge} we obtain Frank’s equation [31]:

$$\forall a, b \in [0, 1] f_{\wedge}(a, b) - f_{\wedge}(1 - a, 1 - b) = a + b + 1$$

Frank [31] showed that for f_{\wedge} to satisfy this equation it must correspond to an ordinal sum of members of the following parameterised family of t-norms:

DEFINITION 5 For parameter $s \in [0, \infty)$

$$f_{\wedge, s}(y_1, y_2) = \begin{cases} \min(y_1, y_2) & : s = 0 \\ \log_s \left(1 + \frac{(s^{y_1} - 1)(s^{y_2} - 1)}{s - 1} \right) & : s > 0, s \neq 1 \\ y_1 \times y_2 & : s = 1 \end{cases}$$

In Frank’s t-norms the parameter s in $f_{\wedge, s}(\chi_{\theta}(x), \chi_{\varphi}(x))$ effectively provides a measure of the dependence between the membership functions of the expressions θ and φ . The smaller the value of s the stronger the dependence (see for example [2] or [46]).

2.2 Functionality and Truth-Functionality

As described in the previous section fuzzy logic [110] is truth-functional, a property which significantly reduces both the complexity and storage requirements of the calculus. Truth-functionality is, however, a rather strong assumption that significantly reduces the number of standard Boolean properties that can be satisfied by the calculus. For instance, Dubois and Prade

[20] effectively showed that no non-trivial truth-functional calculus can satisfy idempotence together with the law of excluded middle. ,

THEOREM 6 *Dubois and Prade [20]*

If χ is truth-functional and satisfies both idempotence and the law of excluded middle then $\forall \theta \in LE \forall x \in \Omega \chi_\theta(x) \in \{0, 1\}$

Proof

From theorem 3 we have that the only idempotent t-conorm is max. Now $\forall \theta \in LE \forall x \in \Omega \chi_{\neg\theta}(x) = f_{\neg}(\chi_\theta(x))$. Hence, by the law of excluded middle $\max(\chi_\theta(x), f_{\neg}(\chi_\theta(x))) = 1$. Now if $\chi_\theta(x) = 1$ then the result is proven. Otherwise if $f_{\neg}(\chi_\theta(x)) = 1$ then by negation axiom N4 $\chi_\theta(x) = f_{\neg}(f_{\neg}(\chi_\theta(x))) = f_{\neg}(1) = 0$ by negation axiom N1 as required.

Elkan [30] somewhat controversially proved a related result for Zadeh's original min / max calculus. Elkan's result focuses on the restrictions imposed on membership values if this calculus is to satisfy a particular classical equivalence relating to re-expressions of logical implication. Clearly this theorem is weaker than that of Dubois and Prade [20] in that it only concerns one particular choice of t-norm, t-conorm and negation function.

THEOREM 7 *Elkan [30]*

Let $f_{\wedge}(a, b) = \min(a, b)$, $f_{\vee}(a, b) = \max(a, b)$ and $f_{\neg}(a) = 1 - a$. For this calculus if $\forall \theta, \varphi \in LE$ and $\forall x \in \Omega \chi_{\neg(\theta \wedge \neg\varphi)}(x) = \chi_{\varphi \vee (\neg\theta \wedge \neg\varphi)}(x)$ then $\forall \theta, \varphi \in LE$ and $\forall x \in \Omega$ either $\chi_\theta(x) = \chi_\varphi(x)$ or $\chi_\varphi(x) = 1 - \chi_\theta(x)$

The controversy associated with this theorem stems mainly from Elkan's assertion in [30] that such a result means that previous successful practical applications of fuzzy logic are somehow paradoxical. The problem with Elkan's attack on fuzzy logic is that it assumes a priori that vague concepts should satisfy a specific standard logical equivalence, namely $\neg(\theta \wedge \neg\varphi) \equiv \varphi \vee (\neg\theta \wedge \neg\varphi)$. No justification is given for the preservation of this law in the case of vague concepts, except that it corresponds to a particular representation of logical implication. Given such an equivalence and assuming the truth-functional min / max calculus of Zadeh then the above reduction theorem (theorem 7) follows trivially.

In their reply to Elkan, Dubois et al. [23] claim that he has confused the notions of epistemic uncertainty and degree of truth. Measures of epistemic uncertainty, they concede, should satisfy the standard Boolean equivalences while degrees of truth need not. In one sense we agree with this point in that Elkan seems to be confusing modelling the uncertainty associated with the object domain Ω with modelling the vagueness of the concepts in the underlying description language LE . On the other hand we do not agree that you can completely separate these two domains. When You make assertions involving

vague concepts then Your intention is to convey information about Ω (a fact recognized in fuzzy set theory by the linking of fuzzy membership functions with possibility distributions [115]). It does not seem reasonable that questions related to this process should be isolated from those relating to the underlying calculus for combining vague concepts. The way in which You conjunctively combine two concepts L_1 and L_2 must be dependent on the information You want to convey about x when You assert ‘ x is $L_1 \wedge L_2$ ’ and the relationship of this information with that conveyed by the two separate assertions ‘ x is L_1 ’ and ‘ x is L_2 ’. Furthermore, it is not enough to merely state that truth-degrees are different from uncertainty and are (or can be) truth-functional. Rather, we claim that the correct approach is to develop an operational semantics for vague concepts and investigate what calculi emerge. Indeed this emphasis on a semantic approach forms the basis of our main object to Elkan’s work. The problem of what equivalences must be satisfied by vague concepts should be investigated within the context of a particular semantics. It is not helpful to merely select such an equivalence largely arbitrarily and then proceed as if the issue had been resolved.

The theme of operation semantics for vague concepts is one that we will return to in a later section of this chapter and throughout this volume. However, for the moment we shall take a different perspective on the result of Dubois and Prade (and to a lesser degree that of Elkan) by noting that it provides an insight into what a strong assumption truth-functionality actually is. We also suggest that truth-functionality is a special case of a somewhat weaker assumption formalizing the following property: Functionality³ assumes that for any sentence $\theta \in LE$ there exists some mechanism by which $\forall x \in \Omega \chi_\theta(x)$ can be determined only from the values $\{\chi_L(x) : L \in LA\}$ (i.e. the membership values of x across the basic labels). This notion seems to capture the underlying intuition that the meaning of compound vague concepts are derived only from the meaning of their component concepts while, as we shall see in the sequel, avoiding the problems highlighted by the theorems of Dubois and Prade and of Elkan.

DEFINITION 8 *The measure ν on $LE \times \Omega$ is said to be functional if $\forall \theta \in LE$ there is function $f_\theta : [0, 1]^n \rightarrow [0, 1]$ such that $\forall x \in \Omega \nu_\theta(x) = f_\theta(\nu_{L_1}(x), \dots, \nu_{L_n}(x))$*

The following example shows that functional measures are not necessarily subject to the triviality result of Dubois and Prade [20].

EXAMPLE 9 *Functional but Non-Truth Functional Calculus*

Let $LA = \{L_1, L_2\}$ and for $\theta \in LE$ let θ_x denote the proposition ‘ x is θ ’. Now let $\nu_\theta(x)$ denote $P(\theta_x)$ where P is a probability measure on the set of propositions $\{\theta_x : \theta \in LE\}$. Suppose then that according to the probability

measure P the propositions $(L_1)_x$ and $(L_2)_x$ are independent for all $x \in \Omega$. In this case ν is a functional but not truth functional measure. For example,

$$\alpha_1 \equiv L_1 \wedge L_2 : \nu_{L_1 \wedge L_2}(x) = \nu_{L_1}(x) \times \nu_{L_2}(x),$$

$$\alpha_2 \equiv L_1 \wedge \neg L_2 : \nu_{L_1 \wedge \neg L_2}(x) = \nu_{L_1}(x) \times (1 - \nu_{L_2}(x))$$

$$\alpha_3 \equiv \neg L_1 \wedge L_2 : \nu_{\neg L_1 \wedge L_2}(x) = (1 - \nu_{L_1}(x)) \times \nu_{L_2}(x)$$

$$\alpha_4 \equiv \neg L_1 \wedge \neg L_2 : \nu_{\neg L_1 \wedge \neg L_2}(x) = (1 - \nu_{L_1}(x)) \times (1 - \nu_{L_2}(x))$$

However, since ν is defined by a probability measure P then

$$\nu_{L_1 \wedge L_1}(x) = \nu_{L_1}(x) \neq \nu_{L_1}(x) \times \nu_{L_1}(x)$$

except when $\nu_{L_1}(x) = 0$ or $\nu_{L_1}(x) = 1$

Clearly, however, $\nu_{L_1 \wedge L_1}(x)$ can be determined directly from $\nu_{L_1}(x)$ and $\nu_{L_2}(x)$ according to the function $f_{L_1 \wedge L_1}(a, b) = a$. Indeed $\nu_\theta(x)$, for any compound expression θ , can be evaluated recursively from $\nu_{L_1}(x)$ and $\nu_{L_2}(x)$ as a unique linear combinations of $\nu_{\alpha_i}(x) : i = 1, \dots, 4$. For instance,

$$\nu_{L_1 \wedge L_1}(x) = \nu_{L_1 \wedge L_2}(x) + \nu_{L_1 \wedge \neg L_2}(x) =$$

$$\nu_{L_1}(x) \times \nu_{L_2}(x) + \nu_{L_1}(x) \times (1 - \nu_{L_2}(x)) = \nu_{L_1}(x)$$

In General

$$\nu_\theta(x) = \sum_{\alpha_i: \alpha_i \rightarrow \theta} \nu_{\alpha_i}(x)$$

Hence, we have that

$$\nu_{\theta \wedge \theta}(x) = \sum_{\alpha_i: \alpha_i \rightarrow \theta \wedge \theta} \nu_{\alpha_i}(x) = \sum_{\alpha_i: \alpha_i \rightarrow \theta} \nu_{\alpha_i}(x) = \nu_\theta(x) \text{ and}$$

$$\nu_{\theta \vee \neg \theta}(x) = \sum_{\alpha_i: \alpha_i \rightarrow \theta \vee \neg \theta} \nu_{\alpha_i}(x) = \sum_{i=1}^4 \nu_{\alpha_i}(x) = 1$$

Clearly then ν satisfies idempotence and the law of excluded middle, and hence functional calculi are not in general subject to the restrictions of Dubois and Prade's theorem [20]

2.3 Operational Semantics for Membership Functions

In [103] Walley proposes a number of properties that any measure should satisfy if it is to provide an effective means of modelling uncertainty in intelligent systems. These include the following interpretability requirement:

'the measure should have a clear interpretation that is sufficiently defined to guide assessment, to understand the conclusions of the system and use them as a basis for action, and to support the rules for combining and updating measures'

Thus according to Walley an operational semantics should not only provide a means of understanding the numerical levels of uncertainty associated with propositions but must also provide some justification for the underlying calculus. In the case of fuzzy logic [110] this means that any interpretation of membership functions should be consistent with truth-functionality. If this turns out not to be the case then it may be fruitful to investigate new calculi for combining imprecise concepts.

In [25] Dubois and Prade suggest three possible semantics for fuzzy logic. One of these is based on the measure of similarity between elements and prototypes of the concept, while two are probabilistic in nature. In this section we shall review all three semantics and discuss their consistency with the truth-functionality assumption of fuzzy logic. We will also describe a semantics based on the risk associated with making an assertion involving vague concepts (Giles [34]).

2.3.1 Prototype Semantics

A direct link between membership functions and similarity measures has been proposed by a number of authors including Ruspini [89] and Dubois and Prade [25], [28]. The basic idea of this semantics is as follows: For any concept θ it is assumed that there is a set of prototypical instances drawn from the universe Ω of which there is no doubt that they satisfy θ . Let \mathcal{P}_θ denote this set of prototypes for θ . It is also assumed that you have some measure of similarity according to which elements of the domain can be compared. Typically this is assumed to be a function $S : \Omega^2 \rightarrow [0, 1]$ satisfying the following properties:

$$S1 \quad \forall x, y \in \Omega \quad S(x, y) = S(y, x)$$

$$S2 \quad \forall x \in \Omega \quad S(x, x) = 1 \quad ^4$$

The membership function for the concept is then defined to be a subjective measure of the similarity between an element x and the closest prototypical element from \mathcal{P}_θ :

$$\forall x \in \Omega \quad \chi_\theta(x) = \sup \{S(x, y) : y \in \mathcal{P}_\theta\}$$

Clearly then if $x \in \mathcal{P}_\theta$ then $\chi_\theta(x) = 1$ and hence if $\mathcal{P}_\theta \neq \emptyset$ then $\sup \{\chi_\theta(x) : x \in \Omega\} = 1$. Also, if $\mathcal{P}_\theta = \emptyset$ then $\forall x \in \Omega \quad \chi_\theta(x) = 0$ and hence according to prototype semantics all non-contradictory concepts have normalised membership functions.

We now consider the type of calculus for membership functions that could be consistent with prototype semantics. Clearly this can be reduced to the problem of deciding what relationships hold between the prototypes of concepts generated as combinations of more fundamental concepts and the prototypes of the component concepts. In other words, what are the relationships between $\mathcal{P}_{-\theta}$ and \mathcal{P}_θ , between $\mathcal{P}_{\theta \wedge \varphi}$ and \mathcal{P}_θ and \mathcal{P}_φ , and between $\mathcal{P}_{\theta \vee \varphi}$ and \mathcal{P}_θ and \mathcal{P}_φ .

For the case of $\neg\theta$ it would seem uncontroversial to assume that $\mathcal{P}_{\neg\theta} \subseteq (\mathcal{P}_\theta)^c$. Clearly, a prototypical *not tall* person cannot also be a prototypical *tall* person. In general, however, it would not seem intuitive to assume that $\mathcal{P}_{\neg\theta} = (\mathcal{P}_\theta)^c$ since, for example, someone who is not prototypically *tall* may not necessarily be prototypically *not tall*.

For conjunctions of concepts one might naively assume that the prototypes for $\theta \wedge \varphi$ might correspond to the intersection $\mathcal{P}_\theta \cap \mathcal{P}_\varphi$. In this case it can easily be seen that:

$$\forall x \in \Omega \chi_{\theta \wedge \varphi}(x) \leq \min(\chi_\theta(x), \chi_\varphi(x))$$

However, on reflection we might wonder whether a typical *tall and medium* person would be either prototypically *tall* or prototypically *medium*. This is essentially the basis of the objection to prototype theory (as based on Zadeh's min-max calculus) raised by Osherson et al. [75]. For example, they note that when considering the concepts *pet* and *fish* then a guppie is much more prototypical of the conjunction *pet fish* than it is of either of the conjuncts. Interestingly when viewed at the membership function level this suggests that the conjunctive combination of membership functions should not be monotonic (as it is for t-norms) since we would intuitively expect guppie to have a higher membership in the extension of *pet fish* than in either of the extensions of *pet* or *fish*.

In the case of disjunctions of concepts it does seem rather more intuitive that $\mathcal{P}_{\theta \vee \varphi} = \mathcal{P}_\theta \cup \mathcal{P}_\varphi$. For example, the prototypical *happy or sad* people might reasonably be thought to be composed of the prototypically *happy* people together with the prototypically *sad* people. In this case we obtain the strict equality:

$$\forall x \in \Omega \chi_{\theta \vee \varphi}(x) = \max(\chi_\theta(x), \chi_\varphi(x))$$

Osherson et al. [75] argue against the use of prototypes to model disjunctions using a counter example based on the concepts *wealth*, *liquidity* and *investment*. The argument presented in [75] assumes that *wealth* corresponds to *liquidity or investment*, however, while there is certainly a relationship between these concepts it is not at all clear that it is a disjunctive one.

EXAMPLE 10 *Suppose the universe Ω is composed of five people:*

$$\Omega = \{Bill(B), Fred(F), Mary(M), Ethel(E), John(J)\}$$

with the following similarity measure S

$S(x, y)$	B	F	M	E	J
B	1	0.8	0.2	0.8	0.7
F	0.8	1	0.5	0.1	0.6
M	0.2	0.5	1	0.9	0.6
E	0.8	0.1	0.9	1	0.3
J	0.7	0.6	0.6	0.3	1

Now let

$$\mathcal{P}_{tall} = \{B, F\}, \mathcal{P}_{medium} = \{M\}$$

$$\mathcal{P}_{\neg tall} = \{J\}, \mathcal{P}_{\neg medium} = \{B\}$$

The membership functions for tall, medium, not tall and not medium are then determined as follows:

$$\chi_{tall}(B) = \chi_{tall}(F) = 1$$

$$\chi_{tall}(M) = \max(S(M, B), S(M, F)) = \max(0.2, 0.5) = 0.5$$

$$\chi_{tall}(E) = \max(S(E, B), S(E, F)) = \max(0.8, 0.1) = 0.8$$

$$\chi_{tall}(J) = \max(S(J, B), S(J, F)) = \max(0.7, 0.6) = 0.7$$

$$\chi_{medium}(B) = S(B, M) = 0.2, \chi_{medium}(F) = S(F, M) = 0.5$$

$$\chi_{medium}(M) = 1, \chi_{medium}(E) = S(E, M) = 0.9$$

$$\chi_{medium}(J) = S(J, M) = 0.6$$

$$\chi_{\neg tall}(B) = S(B, J) = 0.7, \chi_{\neg tall}(F) = S(F, J) = 0.6$$

$$\chi_{\neg tall}(M) = S(M, J) = 0.6, \chi_{\neg tall}(E) = S(E, J) = 0.3$$

$$\chi_{\neg tall}(J) = 1$$

$$\chi_{\neg medium}(B) = 1, \chi_{\neg medium}(F) = S(F, B) = 0.8$$

$$\chi_{\neg medium}(M) = S(M, B) = 0.2, \chi_{\neg medium}(E) = S(E, B) = 0.8$$

$$\chi_{\neg medium}(J) = S(J, B) = 0.7$$

Now $\mathcal{P}_{tall \vee medium} = \mathcal{P}_{tall} \cup \mathcal{P}_{medium} = \{B, F, M\}$ so that:

$$\chi_{tall \vee medium}(B) = 1, \chi_{tall \vee medium}(M) = 1, \chi_{tall \vee medium}(F) = 1$$

$$\chi_{tall \vee medium}(E) = \max(S(E, B), S(E, F), S(E, M))$$

$$= \max(0.8, 0.1, 0.9) = 0.9,$$

$$\chi_{tall \vee medium}(J) = \max(S(J, B), S(J, F), S(J, M))$$

$$= \max(0.7, 0.6, 0.6) = 0.7$$

Clearly then

$$\forall x \in \Omega \chi_{\text{tall} \vee \text{medium}}(x) = \max(\chi_{\text{tall}}(x), \chi_{\text{medium}}(x))$$

Now $\mathcal{P}_{\text{tall}} \cap \mathcal{P}_{\text{medium}} = \emptyset$ and hence if $\mathcal{P}_{\text{tall} \wedge \text{medium}} = \mathcal{P}_{\text{tall}} \cap \mathcal{P}_{\text{medium}}$ then

$$\forall x \in \Omega \chi_{\text{tall} \wedge \text{medium}}(x) = 0$$

Alternatively, we might expect that Ethel would be a prototypical $\text{tall} \wedge \text{medium}$ person since she has high membership both in tall and medium. Taking

$\mathcal{P}_{\text{tall} \wedge \text{medium}} = \{E\}$ gives:

$$\chi_{\text{tall} \wedge \text{medium}}(B) = S(B, E) = 0.8, \chi_{\text{tall} \wedge \text{medium}}(F) = S(F, E) = 0.1,$$

$$\chi_{\text{tall} \wedge \text{medium}}(M) = S(M, E) = 0.9, \chi_{\text{tall} \wedge \text{medium}}(E) = 1,$$

$$\chi_{\text{tall} \wedge \text{medium}}(J) = S(J, E) = 0.3$$

What seems clear from the above discussion and example is that membership functions based on similarity measures are almost certainly not truth-functional and probably not even functional. For instance, the precise relationship between $\mathcal{P}_{\neg\theta}$ and \mathcal{P}_{θ} and between $\mathcal{P}_{\theta \wedge \varphi}$ and \mathcal{P}_{θ} and \mathcal{P}_{φ} is problem specific. In other words, such relationships are strongly dependent on the meanings of θ and φ . It is the case, however, that if we take the prototypes of disjunctions as corresponding to the union of prototypes of the disjuncts then the resulting calculus will be partially-functional.

The fundamental problem with similarity semantics, however, is not principally related to the functionality of the emergent calculus. Rather it lies with the notion of similarity itself. In some concepts it may be straightforward to define the level of similarity between objects and prototypes. For example, in the case of the concept *tall* we might reasonably measure the similarity between individuals in terms of some monotonically decreasing function of the Euclidean distance between their heights. For other concepts, however, the exact nature of the underlying similarity measure would seem much harder to identify. For instance, consider the concept ‘tree’. Now supposing we could identify a set of prototypical trees, itself a difficult task, how could we quantify the similarity between a variety of different plants and the elements of this prototype set? It is hard to identify an easily measurable attribute of plants that could be used to measure the degree of ‘treeness’. To put it bluntly, the degree of similarity would seem as hard to define as the degree of membership itself!

2.3.2 Risk/Betting Semantics

Risk or betting semantics was proposed by Giles in a series of papers including [34] and [35]. In this semantics the fundamental idea (as described in [34]) is that the membership $\chi_{\theta}(x)$ quantifies the level of risk You are taking

when You assert that ‘ x is θ ’. Formally, let θ_x denote the proposition ‘ x is θ ’ then by asserting θ_x You are effectively saying that You will pay an opponent 1 unit if θ_x is false. Given this gamble we may then suppose that You associate a risk value $\langle \theta_x \rangle \in [0, 1]$ with the assertion θ_x . Such values are likely to be subjective probabilities and vary between agents, however, if $\langle \theta_x \rangle = 0$ we may reasonably suppose that You will be willing to assert θ_x and if $\langle \theta_x \rangle = 1$ You will certainly be unwilling to make such an assertion. We then define Your membership degree of x in θ as:

$$\chi_{\theta}(x) = 1 - \langle \theta_x \rangle$$

this being the subjective probability that θ_x is true. The standard connectives are then interpreted as follows:

Negation: You asserting $\neg\theta_x$ means that You will pay your opponent 1 unit if they will assert θ_x .

In this case You will be willing to assert $\neg\theta_x$ if You are sufficiently sure that θ_x is false, since in this case You will be repaid Your 1 unit by Your opponent. Hence, Your risk when asserting $\neg\theta_x$ is equivalent to one minus Your risk when asserting θ_x . Correspondingly,

$$\chi_{\neg\theta}(x) = 1 - \chi_{\theta}(x)$$

Conjunction: You asserting $(\theta \wedge \varphi)_x$ means that You agree to assert θ_x or φ_x , where the choice is made by Your opponent.

In this case the risk of asserting $(\theta \wedge \varphi)_x$ is the maximum of the risks associated with asserting θ_x and φ_x , since You have no way of knowing which of these two options Your opponent will choose. Hence, given the defined relationship between membership and risk we have

$$\chi_{\theta \wedge \varphi}(x) = \max(\chi_{\theta}(x), \chi_{\varphi}(x))$$

Disjunction: You asserting $(\theta \vee \varphi)_x$ means that You agree to assert either θ_x or φ_x , where the choice is made by You.

In this case, since You have the choice of which of the two statements to assert it is rational for You to choose the statement with minimum associated risk. In other words, $\langle (\theta \vee \varphi)_x \rangle$ should correspond to $\min(\langle \theta_x \rangle, \langle \varphi_x \rangle)$. Hence, the corresponding membership degree will satisfy:

$$\chi_{\theta \vee \varphi}(x) = \min(\chi_{\theta}(x), \chi_{\varphi}(x))$$

Although this semantics captures the fuzzy set calculus proposed by Zadeh [110], there is something odd about the truth-functional way in which the risks associated with asserting a compound expression are calculated. For instance,

the rule for evaluating the risk of conjunction seems to implicitly assume that there is no logical relationship between the conjuncts. To see this consider the contradictory assertion $(\theta \wedge \neg\theta)_x$. In betting semantics this means that You are willing to either pay one unit to your opponent if θ_x is false (bet 1), or to pay Your opponent one unit if they will assert θ_x (bet 2). The choice between these two bets is then made by Your opponent. The main problem with this assertion is that given an informed opponent You are certain to lose. For instance, assuming Your opponent knows the truth value of θ_x then they would choose between bets 1 and 2 accordingly, as follows: If θ_x is true then Your opponent will pick bet 2 and You will lose 1 unit. Alternatively, if θ_x is false then Your opponent will pick bet 1 and You will lose 1 unit. Hence, given an informed opponent with knowledge of the truth value of θ_x , You are certain to lose 1 unit. In fact even when there is some doubt regarding the judgement of Your opponent You will tend to lose on average. For instance, if p is the probability that Your opponent will be correct regarding the truth value of θ_x then Your expected loss will be $p \times 1 + (1 - p) \times 0 = p$. This suggests that Your actual risk when asserting $(\theta \wedge \neg\theta)_x$ is directly related to p and relatively independent of $\langle\theta_x\rangle$. Furthermore, since You will expect to lose in all cases except when $p = 0$, this being when Your opponent is definitely going to be wrong regarding θ_x , it is hard to imagine a scenario in which You would ever assert such a contradiction.

This problem with assuming a truth-functional calculus seems to have been recognised by Giles in later work [35] where he related the degree of membership $\chi_\theta(x)$ to a more general utility function $h_\theta(x)$. The latter is intended to quantify the degree of utility, possibly negative, which You will receive on asserting θ_x . There is no reason, however, why such a utility measure should be truth-functional, a fact highlighted by Giles [35] who comments that his research suggests ‘that there is no viable truth-functional representation for conjunction and disjunction of fuzzy sentences’.

2.3.3 Probabilistic Semantics

Of the three semantics proposed by Dubois and Prade [25] two are probabilistic in nature. However, they do not offer a naive interpretation of fuzzy membership functions by, for example, claiming that they are simply probability distributions or density function that have not been normalised (e.g. Lavolette et al. [58] propose modelling linguistic concepts such as ‘medium’ and ‘fast’ using probability density functions). Rather they aim to model vague concepts in terms of the underlying uncertainty or variance associated with their meaning. In this section we discuss three such probabilistic semantics. This will include an in depth examination of random set semantics as mentioned by Dubois and Prade in [25] and developed at length by Goodman and Nguyen in a number of articles including [36], [37], [38], [72] and [73]. This work will be related to

the voting model ([9], [32] and [59]) and the context model ([33] and [47]). We would ask for the reader's indulgence with respect to this extended discussion of random set interpretations of membership functions on the grounds that it is closely related to the semantics proposed subsequently in this volume. In addition, we will outline the likelihood interpretation of fuzzy sets as suggested by Hisdal [44] and show that, as indicated by Dubois and Prade [25], it is strongly related to the random set approach.

2.3.3.1 Random Set Semantics

In essence random sets are set valued variables and hence can be defined in terms of a mapping between two measurable domains as follows:

DEFINITION 11 *Let \mathcal{B} be a σ -algebra on a universe U_1 and let \mathcal{A} be a σ -algebra on 2^{U_2} , the power set of a second universe U_2 . Then R is a random set from $\langle U_1, \mathcal{B} \rangle$ into $\langle 2^{U_2}, \mathcal{A} \rangle$ if $R : U_1 \rightarrow 2^{U_2}$ is a $\mathcal{B} - \mathcal{A}$ measurable function.*

If U_2 is finite and P is a probability measure on \mathcal{B} then we can define a mass assignment (i.e. a probability distribution for R) according to:

$$\forall T \subseteq U_2 \quad m(T) = P(R = T) = P(\{z \in U_1 : R(z) = T\})$$

This then generates a probability measure \mathcal{M} on \mathcal{A} as follows:

$$\forall A \in \mathcal{A} \quad \mathcal{M}(A) = P(R \in A) = \sum_{T \in A} m(T)$$

DEFINITION 12 *Let R be a random set into 2^{U_2} then the fixed (or single) point coverage function is a function $cf_R : U_2 \rightarrow [0, 1]$ such that:*

$$\forall w \in U_2 \quad cf_R(w) = P(w \in R) = \sum_{T \subseteq U_2 : w \in T} m(T)$$

EXAMPLE 13 *Let $U_1 = \{a_1, a_2, a_3, a_4, a_5\}$ and $U_2 = \{b_1, b_2, b_3\}$, and let $\mathcal{B} = 2^{U_1}$ and $\mathcal{A} = 2^{2^{U_2}}$. Also, let \mathcal{P} be defined according to the following values on the singleton sets:*

$$P(a_1) = 0.1, P(a_2) = 0.4, P(a_3) = 0.2, P(a_4) = 0.1, P(a_5) = 0.2$$

A possible random set from $\langle U_1, \mathcal{B} \rangle$ into $\langle 2^{U_2}, \mathcal{A} \rangle$ is:

$$R(a_1) = \{b_1, b_2\}, R(a_2) = \{b_2, b_3\}, R(a_3) = \{b_1, b_2, b_3\}, \\ R(a_4) = \{b_1, b_2, b_3\}, R(a_5) = \{b_2, b_3\}$$

The corresponding mass assignment is given by:

$$\begin{aligned} m(\{b_1, b_2\}) &= P(a_1) = 0.1, \quad m(\{b_2, b_3\}) = P(a_2) + P(a_5) \\ &= 0.4 + 0.2 = 0.6, \quad m(\{b_1, b_2, b_3\}) = P(a_3) + P(a_4) = 0.2 + 0.1 = 0.3 \end{aligned}$$

The fixed point coverage function can then be evaluated as follows:

$$\begin{aligned} cf_R(b_1) &= m(\{b_1, b_2\}) + m(\{b_1, b_2, b_3\}) = 0.1 + 0.3 = 0.4 \\ cf_R(b_2) &= m(\{b_1, b_2\}) + m(\{b_2, b_3\}) + m(\{b_1, b_2, b_3\}) = 1 \\ cf_R(b_3) &= m(\{b_2, b_3\}) + m(\{b_1, b_2, b_3\}) = 0.6 + 0.3 = 0.9 \end{aligned}$$

It is interesting to note that although any given mass assignment on 2^{U_2} yields a unique coverage function the converse does not hold. That is, there is generally a (sometimes infinite) set of mass assignments with the same fixed point coverage function. For instance, consider the coverage function given in example 13. What other mass assignments also have this coverage function? Since $cf_R(b_2) = 1$ it follows that, for any such mass assignment, $m(T) = 0$ for any subset T of U_2 not containing b_2 . Now let:

$$\begin{aligned} m(\{b_1, b_2, b_3\}) &= m_1, \quad m(\{b_1, b_2\}) = m_2, \quad m(\{b_2, b_3\}) = m_3 \\ \text{and } m(\{b_2\}) &= m_4 \end{aligned}$$

Now from the equation for the fixed point coverage function given in definition 12 we have that:

$$\begin{aligned} m_1 + m_2 + m_3 + m_4 &= 1, \quad m_1 + m_2 = 0.4, \quad m_1 + m_3 = 0.9 \text{ so that} \\ m_2 &= 0.4 - m_1, \quad m_3 = 0.9 - m_1 \text{ and } m_4 = m_1 - 0.3 \text{ where } m_1 \in [0.3, 0.4] \end{aligned}$$

Clearly, the mass assignment in example 13 corresponds to the case where $m_1 = 0.3$. Another, interesting case is where $m_1 = 0.4$ giving the mass assignment:

$$m(\{b_1, b_2, b_3\}) = 0.4, \quad m(\{b_2, b_3\}) = 0.5, \quad m(\{b_2\}) = 0.1$$

In this case where the subsets with non-zero mass form a nested hierarchy the underlying random set is referred to as consonant:

DEFINITION 14 A consonant random set R into 2^{U_2} is such that

$$\bigcup_{a \in U_1} R(a) = \{G_1, \dots, G_t\} \text{ where } G_1 \subset G_2 \subset \dots \subset G_t \subseteq U_2$$

In fact, as the following theorem shows, there is a unique consonant mass assignment consistent with any particular coverage function.

THEOREM 15 Given a fixed point coverage function cf_R for which $\{cf_R(b) : b \in U_2\} = \{y_1, \dots, y_n\}$ ordered such that $y_i > y_{i+1} : i = 1, \dots, n-1$ then any consonant random set with this coverage function must have the following mass assignment:

For $F_i = \{b \in U_2 : cf_R(b) \geq y_i\}$

$m(F_n) = y_n, m(F_i) = y_i - y_{i+1} : i = 1, \dots, n-1$ and $m(\emptyset) = 1 - y_1$

Proof

Since R is consonant we have that

$\{T : m(T) > 0\} = \{G_1, \dots, G_t\}$ where $G_1 \subset G_2 \subset \dots \subset G_t \subseteq U_2$

Let $\bar{b}_i = \{b \in U_2 : cf_R(b) = y_i\}$ then there is a minimal value $i^* \in \{1, \dots, t\}$ such that $\bar{b}_i \subseteq G_{i^*}$ and $\bar{b}_i \cap G_j = \emptyset$ for $j < i^*$. Hence,

$$y_i = \sum_{k=i^*}^t m(G_k)$$

Now if $j < i$ then $\bar{b}_j \subseteq G_{i^*}$ since otherwise $j^* > i^*$ so that

$$y_j = \sum_{k=j^*}^t m(G_k) < \sum_{k=i^*}^t m(G_k) = y_i$$

which is a contradiction. Also for $j > i$ $\bar{b}_j \cap G_{i^*} = \emptyset$ since otherwise $j^* < i^*$ so that

$$y_j = \sum_{k=j^*}^t m(G_k) > \sum_{k=i^*}^t m(G_k) = y_i$$

which is a contradiction. From this it follows that:

$$G_{i^*} = \bigcup_{k=1}^i \bar{b}_k = \{b \in U_2 : cf_R(b) \geq y_i\}$$

Now it can easily be seen that $(i+1)^* = i^* + 1$ hence

$$m(G_{i^*}) = \sum_{k=i^*}^t m(G_k) - \sum_{k=i^*+1}^t m(G_k) = y_i - y_{i+1} \text{ for } i = 1, \dots, n-1$$

Now $G_{n^*} = \{b \in U_2 : cf_R(b) > 0\}$ and therefore $n^* = t$. From this it follows that

$$y_n = m(G_t)$$

Also, suppose that $j < 1^*$ then either $G_j = \emptyset$ or $\exists b \in G_j$ such that $b \notin G_{1^*}$ which implies that $cf_R(b) > y_1$. This is a contradiction, so either $1^* = 1$ or $1^* = 2$ and $G_1 = \emptyset$. Hence, w.l.o.g and by relabelling the relevant sets we may assume that:

$$\{T : m(T) > 0\} = \{\emptyset, F_1, \dots, F_n\} \text{ where } F_i = G_{i^*} : i = 1, \dots, n$$

Finally, as required

$$m(\emptyset) = 1 - \sum_{k=1}^n m(F_k) = 1 - \left(\sum_{k=1}^{n-1} y_k - y_{k+1} \right) - y_n = 1 - y_1 \quad \square$$

Another interesting mass assignment sharing the same coverage function as given in example 13 is obtained by setting $m_1 = 0.4 \times 0.9 \times 1 = 0.36$. This is the solution with maximum entropy and has the general form:

$$\forall T \subseteq U_2 \quad m(T) = \left(\prod_{b \in T} cf_R(b) \right) \times \left(\prod_{b \notin T} (1 - cf_R(b)) \right)$$

Essentially, the idea of random set semantics is that vague concepts are concepts for which the exact definition is uncertain. In this case the extension of a vague concept θ is defined by a random set R_θ into 2^Ω (i.e. $U_2 = \Omega$) with associated mass assignment m . The membership function $\chi_\theta(x)$ is then taken as coinciding with the fixed point coverage function of R_θ so that:

$$\forall x \in \Omega \quad \chi_\theta(x) = cf_{R_\theta}(x) = P(x \in R_\theta) = \sum_{T \subseteq \Omega: x \in T} m(T)$$

Within the context of random set semantics, given that for a particular element of U_1 the extension of a concept is a crisp set of elements from Ω , it is perhaps natural to assume that the following classical laws hold when combining vague concepts:

$$\begin{aligned} \forall \theta \in LE \quad \forall a \in U_1 \quad R_{\neg\theta}(a) &= R_\theta(a)^c \\ \forall \theta, \varphi \in LE \quad \forall a \in U_1 \quad R_{\theta \wedge \varphi}(a) &= R_\theta(a) \cap R_\varphi(a) \\ \forall \theta, \varphi \in LE \quad \forall a \in U_1 \quad R_{\theta \vee \varphi}(a) &= R_\theta(a) \cup R_\varphi(a) \end{aligned}$$

In this case it can easily be seen that the associated calculus of membership (fixed point coverage) functions will, in general, be neither functional or truth-functional. Certainly such a calculus would satisfy the standard Boolean laws including idempotence (since $R_\theta(a) \cap R_\theta(a) = R_\theta(a)$), the law of excluded middle (since $R_\theta(a) \cup R_\theta(a)^c = \Omega$) and the law of non-contradiction (since $R_\theta(a) \cap R_\theta(a)^c = \emptyset$).

2.3.3.2 Voting and Context Model Semantics

One aspect of the random set semantics that remains unclear from the above discussion is the exact nature of the uncertainty regarding the extension of θ . In more formal terms this corresponds to asking what is the exact nature of the universe U_1 . One possibility is to assume that the random set uncertainty comes from the variation in the way that concepts are defined across a population. That is we take U_1 to be a population of individuals V each of whom are asked to provide an exact (crisp) extension of θ . Alternatively, for finite universes with small numbers of elements, these extensions may be determined implicitly by asking each individual whether or not they think that each element satisfies θ . This is the essence of the voting model for fuzzy sets proposed originally by Black [9] and later by Gaines [32]. Hence, in accordance with the random set model of fuzzy membership we have that

$$\chi_\theta(x) = P(\{v \in V : x \in R_\theta(v)\})$$

and assuming P is the uniform distribution then

$$= \frac{|\{v \in V : x \in R_\theta(v)\}|}{|V|}$$

Hence, when P is the uniform distribution on voters then $\chi_\theta(x)$ is simply the proportion of voters that include x in the extension of θ (or alternatively agree that 'x is θ '). Now assuming that voters apply the classical rules for the logical connectives then as discussed above such a random set based model will not be truth-functional. In fact it is well known that such measures of uncertainty defined in terms of relative frequency are probability measures (see Paris [78] for an exposition). As an alternative, Lawry [59] proposed a non-classical mechanism according to which voters could decide whether or not an element satisfied a concept as defined in terms of a logical combinations of labels. This mechanism corresponds to the following non-standard extension of the classical logic notion of a valuation.

DEFINITION 16 *A fuzzy valuation for instance $x \in \Omega$ is a function $F_x : LE \times [0, 1] \rightarrow \{1, 0\}$ satisfying the following conditions (see figure 2.3):*

- (i) $\forall \theta \in LE$ and $\forall y, y' \in [0, 1]$ such that $y \leq y'$ then $F_x(\theta, y) = 0 \Rightarrow F_x(\theta, y') = 0$
- (ii) $\forall \theta, \varphi \in LE$ and $\forall y \in [0, 1]$ $F_x(\theta \wedge \varphi, y) = 1$ iff $F_x(\theta, y) = 1$ and $F_x(\varphi, y) = 1$ (see figure 2.4)
- (iii) $\forall \theta, \varphi \in LE$ and $\forall y \in [0, 1]$ $F_x(\theta \vee \varphi, y) = 1$ iff $F_x(\theta, y) = 1$ or $F_x(\varphi, y) = 1$
- (iv) $\forall \theta \in LE$ and $\forall y \in [0, 1]$ $F_x(\neg\theta, y) = 1$ iff $F_x(\theta, 1 - y) = 0$ (see figure 2.5)

Fuzzy valuations are an extension of classical valuations where the truth value assigned is dependent not only on the expression but also on a parameter y between 0 and 1 representing the degree of scepticism of the voter. The closer y is to 0 the less sceptical the voter and the more likely they are to be convinced of the truth of any given statement (i.e. to assign a truth-value of 1) and conversely the closer y is to 1 the more sceptical the voter and the more likely it is that they will not be convinced of the truth of the expression (i.e. to assign a truth-value of 0). The scepticism level, then, should be thought of as representing an internal state of the agent according to which their behaviour is more or less cautious. Fuzzy valuations attempt to capture formally the following description of voter behaviour as given by Gaines [32]:

‘members of the population each evaluated the question according to the same criteria but applied a different threshold to the resulting evidence, or ‘feeling’. The member with the lowest threshold would then always respond with a yes answer when any other member did, and so on up the scale of thresholds.’

Now assuming y varies across voters we may suppose that the probability distribution on scepticism levels is given by a probability measure ρ on the Borel subsets of $[0, 1]$. Given this measure, the membership degree $\chi_\theta(x)$ relative to a fuzzy valuation F_x can then be defined as the probability of that a voter has a scepticism level y such that $F_x(\theta, y) = 1$. More formally,

DEFINITION 17 *Given a fuzzy valuation F_x we can define a corresponding membership function for any $\theta \in LE$ according to:*

$$\chi_\theta(x) = \rho(\{y \in [0, 1] : F_x(\theta, y) = 1\})$$

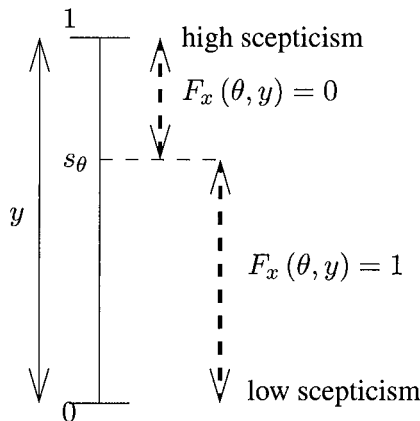


Figure 2.3: Diagram showing how fuzzy valuation F_x varies with scepticism level y

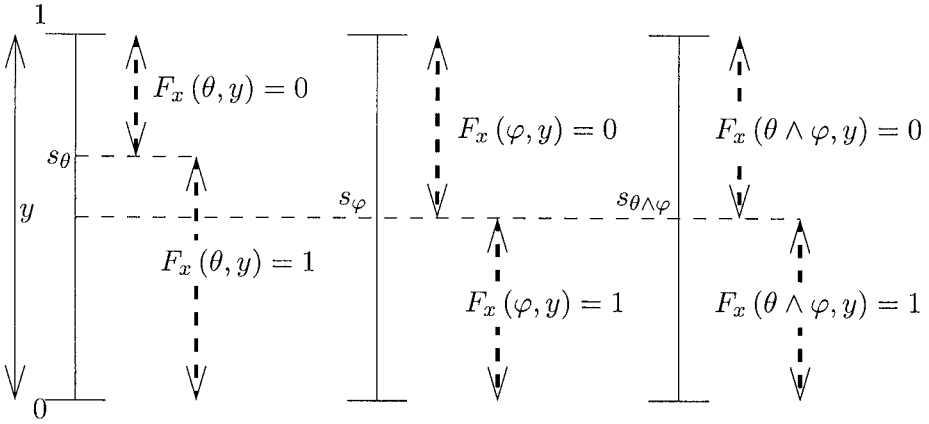


Figure 2.4: Diagram showing the rule for evaluating the fuzzy valuation of a conjunction at varying levels of scepticism

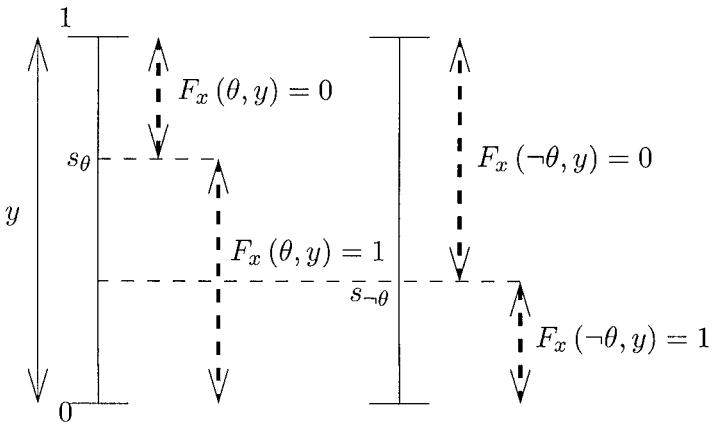


Figure 2.5: Diagram showing the rule for evaluating the fuzzy valuation of a negation at varying levels of scepticism

THEOREM 18 For membership degrees based on fuzzy valuation F_x we have that

$$\forall \theta, \varphi \in LE \chi_{\theta \wedge \varphi}(x) = \min(\chi_{\theta}(x), \chi_{\varphi}(x))$$

Proof

$$\begin{aligned} \chi_{\theta \wedge \varphi}(x) &= \rho(\{y \in [0, 1] : F_x(\theta, y) = 1, F_x(\varphi, y) = 1\}) \text{ and} \\ &\{y \in [0, 1] : F_x(\theta, y) = 1, F_x(\varphi, y) = 1\} = \\ &\{y \in [0, 1] : F_x(\theta, y) = 1\} \cap \{y \in [0, 1] : F_x(\varphi, y) = 1\} \end{aligned}$$

Now $\forall \theta \in LE$ let

$s_\theta = \sup \{y \in [0, 1] : F_x(\theta, y) = 1\}$ so that either
 $\{y \in [0, 1] : F_x(\theta, y) = 1\} = [0, s_\theta]$ or $= [0, s_\theta)$

Then w.l.o.g suppose that $\chi_\varphi(x) \leq \chi_\theta(x)$ which implies that $s_\varphi \leq s_\theta$. In this case (see figure 2.4),

$\{y \in [0, 1] : F_x(\theta, y) = 1\} \cap \{y \in [0, 1] : F_x(\varphi, y) = 1\} =$ either
 $[0, s_\theta] \cap [0, s_\varphi] = [0, s_\varphi]$ or $[0, s_\theta] \cap [0, s_\varphi) = [0, s_\varphi)$ or
 $[0, s_\theta) \cap [0, s_\varphi] = [0, s_\varphi]$ or $[0, s_\theta) \cap [0, s_\varphi) = [0, s_\varphi)$

In all of these cases

$\{y \in [0, 1] : F_x(\theta, y) = 1, F_x(\varphi, y) = 1\} = \{y \in [0, 1] : F_x(\varphi, y) = 1\}$

and hence $\chi_{\theta \wedge \varphi}(x) = \chi_\varphi(x)$ as required. \square

THEOREM 19 For membership degrees based on fuzzy valuation F_x we have that

$\forall \theta, \varphi \in LE \chi_{\theta \vee \varphi}(x) = \max(\chi_\theta(x), \chi_\varphi(x))$

Proof

Follows similar lines to theorem 18. \square

THEOREM 20 Provided that ρ is a symmetric probability measure satisfying $\forall a, b \in [0, 1] \rho([a, b]) = \rho([1 - b, 1 - a])$ then for membership degrees based on fuzzy valuation F_x we have that

$\forall \theta \in LE \chi_{\neg \theta}(x) = 1 - \chi_\theta(x)$

Proof

By the symmetry condition on ρ it follows that $\forall a \in [0, 1] \rho([0, a]) = \rho([1 - a, 1])$ and $\rho([0, a)) = \rho((1 - a, 1])$. Also recall that either $\{y \in [0, 1] : F_x(\theta, y) = 1\} = [0, s_\theta]$ or $= [0, s_\theta)$.

Now

$\chi_{\neg \theta}(x) = \rho(\{y \in [0, 1] : F_x(\neg \theta, y) = 1\}) =$ by definition 16 part (iv)
 $\rho(\{y \in [0, 1] : F_x(\theta, 1 - y) = 0\})$

Supposing that

$\{y \in [0, 1] : F_x(\theta, y) = 1\} = [0, s_\theta]$ then by definition 16 part (iv)
 $F_x(\neg \theta, y) = 1$ iff $1 - y > s_\theta$ iff $y < 1 - s_\theta$ and therefore (see figure 2.5)
 $\{y \in [0, 1] : F_x(\neg \theta, y) = 1\} = [0, 1 - s_\theta) = [1 - s_\theta, 1]^c$ therefore
 $\rho(\{y \in [0, 1] : F_x(\neg \theta, y) = 1\}) = 1 - \rho([1 - s_\theta, 1]) = 1 - \rho([0, s_\theta])$
 $= 1 - \chi_\theta(x)$ by the symmetry condition on ρ

Alternatively supposing that

$$\{y \in [0, 1] : F_x(\theta, y) = 1\} = [0, s_\theta] \text{ then}$$

$$\{y \in [0, 1] : F_x(\neg\theta, y) = 1\} = [0, 1 - s_\theta] = (1 - s_\theta, 1]^c \text{ and therefore}$$

$$\rho(\{y \in [0, 1] : F_x(\neg\theta, y) = 1\}) = 1 - \rho((1 - s_\theta, 1]) = 1 - \rho([0, s_\theta])$$

$$= 1 - \chi_\theta(x) \quad \square$$

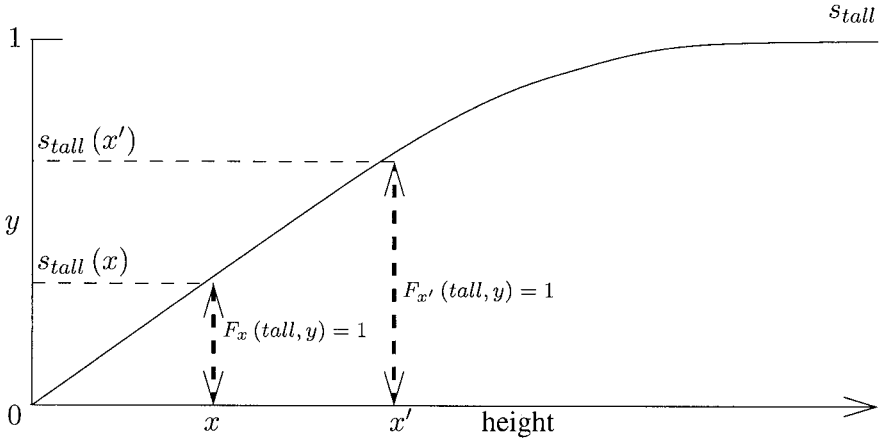


Figure 2.6: Diagram showing how the range of scepticism values for which an individual is considered tall increases with height

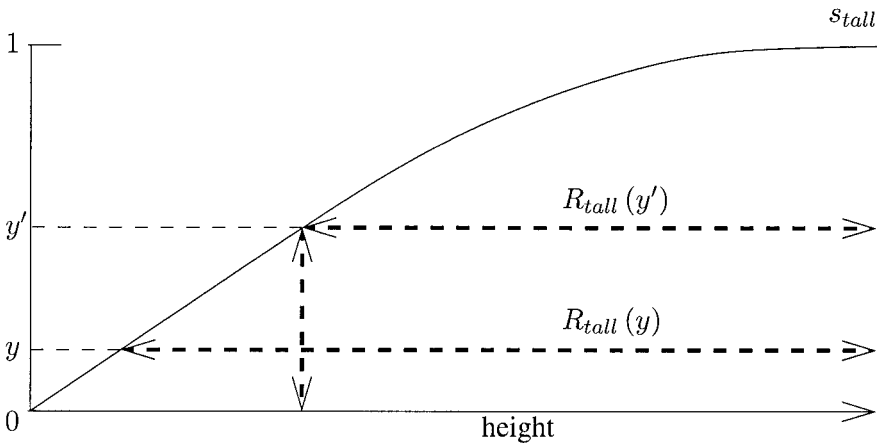


Figure 2.7: Diagram showing how the extension of tall varies with the y

Now as for each $x \in \Omega$ it is assumed that every voter defines a distinct fuzzy valuation F_x across which the upper bound on scepticism levels for which θ is true, s_θ , will vary. Hence, we can view $s_\theta : \Omega \rightarrow [0, 1]$ as a function of x .

For example, we might expect the value of s_{tall} to be an increasing function of height since the greater the height of an individual the more likely a voter will be to agree that this individual is *tall* even if that voter has a relatively high scepticism level (see figure 2.6). Also, given such a set of fuzzy valuations then for each level y we can naturally generate the extension of a concept θ by considering all elements x for which $F_x(\theta, y) = 1$ (see figure 2.7). This motivates the following definition:

DEFINITION 21 *Given a set of fuzzy valuations F_x for every $x \in \Omega$ then for any $\theta \in LA$ the extension of θ is naturally defined by the function $R_\theta : [0, 1] \rightarrow 2^{U_2}$ as follows:*

$$\forall y \in [0, 1] \quad R_\theta(y) = \{x \in \Omega : F_x(\theta, y) = 1\}$$

The extension of a concept defined in this way, relative to a set of fuzzy valuations, corresponds to a consonant random set as shown by the following theorem.

THEOREM 22 *For any $\theta \in LE$, given a set of fuzzy valuations F_x for every $x \in \Omega$, then R_θ as defined in definition 21 is consonant random set from $\langle [0, 1], \mathcal{B} \rangle$ into $\langle 2^\Omega, 2^{2^\Omega} \rangle$ where \mathcal{B} are the Borel subsets of $[0, 1]$ and (for mathematical simplicity) it is assumed that Ω is finite.*

Proof

For $T \subseteq \Omega$ $R_\theta^{-1}(T) = \{y \in [0, 1] : R_\theta(y) = T\}$ can be determined as follows: Let

$$\bar{s}_T = \min(s_\theta(x) : x \in T) \quad \text{and} \quad \underline{s}_T = \max(s_\theta(x) : x \notin T)$$

Now if $y \in R_\theta^{-1}(T)$ then $y \leq \bar{s}_T$ since otherwise $\exists x \in T$ such that $y > s_\theta(x)$ which implies $F_x(\theta, y) = 0$ and hence $x \notin R_\theta(y)$ which would mean that $R_\theta(y) \neq T$. Similarly, if $y \in R_\theta^{-1}(T)$ then $y \geq \underline{s}_T$ since otherwise $\exists x \in T^c$ such that $y < s_\theta(x)$ which implies $F_x(\theta, y) = 1$ and hence $x \in R_\theta(y)$ which would mean that $R_\theta(y) \neq T$. Also, if $y < \bar{s}_T$ and $y > \underline{s}_T$ then $\forall x \in T$ $F_x(\theta, y) = 1$ and $\forall x \in T^c$ $F_x(\theta, y) = 0$ so that $R_\theta(y) = T$.

From this we can see that

$$R_\theta^{-1}(T) = \text{either} \\ \emptyset \text{ (if } \bar{s}_T < \underline{s}_T \text{) or } [\underline{s}_T, \bar{s}_T] \text{ or } [\underline{s}_T, \bar{s}_T) \text{ or } (\underline{s}_T, \bar{s}_T] \text{ or } (\underline{s}_T, \bar{s}_T)$$

All of these sets are Borel measurable as required.

We now show that R_θ is consonant. If $y' > y$ then by definition 16 part (i) it

follows that

$\{x \in \Omega : F_x(\theta, y) = 0\} \subseteq \{x \in \Omega : F_x(\theta, y') = 0\}$ and therefore
 $\{x \in \Omega : F_x(\theta, y') = 0\}^c \subseteq \{x \in \Omega : F_x(\theta, y) = 0\}^c$ which implies
 $\{x \in \Omega : F_x(\theta, y') = 1\} \subseteq \{x \in \Omega : F_x(\theta, y) = 1\}$ and hence that
 $R_\theta(y') \subseteq R_\theta(y)$ as required. \square

THEOREM 23 Given R_θ as defined in definition 21 and χ_θ as defined in definition 17 then

$$\forall x \in \Omega \chi_\theta(x) = cf_{R_\theta}(x)$$

Proof

$$\begin{aligned} \forall x \in \Omega cf_{R_\theta}(x) &= \sum_{S:x \in S} m(S) = \sum_{S:x \in S} \rho(\{y \in [0, 1] : R_\theta(y) = S\}) \\ &= \rho(\{y \in [0, 1] : x \in R_\theta(y)\}) = \rho(\{y \in [0, 1] : F_x(\theta, y) = 1\}) = \chi_\theta(x) \quad \square \end{aligned}$$

While fuzzy valuations provide an interesting mechanism according to which a truth-functional calculus can emerge from a random set interpretation of membership functions the theory remains problematic with respect to its treatment of negation. Specifically the negation rule given in definition 16 part (iv) is hard to justify. Lawry [59] suggests that:

‘In order to determine a truth value for $\neg\theta$ while in state y the agent converts to a dual state $1 - y$ to evaluate the truth value of θ . The truth value of $\neg\theta$ is then taken to be the opposite of this truth value for θ .’

This, however, would appear to be a somewhat convoluted method for evaluating the negation of an expression without a clear semantic justification. Paris [80] proposes the following alternative justification:

‘if voters with a low degree of scepticism y reject θ (i.e. $F(\theta, y) = 0$ for some low value of y) then other voters, even those with a relatively high degree of scepticism, would see this as support for $\neg\theta$ and be influenced to vote accordingly’

The problem with this argument is that it presupposes that individual voters will have access not only to the voting response of other voters but also their scepticism level at the time of voting. This would seem a very unrealistic assumption.

In [33] Gebhardt and Kruse proposed a variant on the voting model in which the elements of U_1 are interpreted as different contexts across which vague concepts have different extensions. For example, in [47] it is suggested that in the case where $LA = \{very\ short, short, medium, \dots\}$ the contexts (i.e. elements of U_1) might correspond to nationalities such as Japanese, American, Swede, etc. Once again, however, it is clear that such a calculus will not in

general be functional and as stated in [47] it is only possible to restrict $\chi_{\theta \wedge \varphi}(x)$ to the standard probabilistic interval so that:

$$\forall x \in \Omega, \forall \theta, \varphi \in LE$$

$$\chi_{\theta \wedge \varphi}(x) \in [\max(0, \chi_{\theta}(x) + \chi_{\varphi}(x) - 1), \min(\chi_{\theta}(x), \chi_{\varphi}(x))]$$

Interestingly, [47] indentify a special case of the context model for which a limited form of functionality does exist. Here it is supposed that the total set of labels is comprised of two distinct subsets LA_1 and LA_2 with different sets of contexts (C_1 and C_2 respectively) so that $U_1 = C_1 \times C_2$. For example, LA_1 might be the set of height labels described above with C_1 the associated set of nationality contexts, while LA_2 might be a set of labels relating to income (e.g. *high, low*) and C_2 a set of residential areas. Now it is reasonable to assume that for such different sets of contexts the occurrence of a particular context from C_1 will be independent of the occurrence of any context from C_2 . Hence, if P_1 is the probability distribution on C_1 and P_2 is the probability distribution on C_2 then the probability distribution on the joint space of contexts U_1 will be $P_1 \times P_2$. In this case, if LE_1 is the set of expression generated from labels LA_1 and LE_2 is the set of expression generated from labels LA_2 then for $\theta \in LE_1$ and $\varphi \in LE_2$ is can easily be seen that:

$$\forall x \in \Omega \chi_{\theta \wedge \varphi}(x) = \chi_{\theta}(x) \times \chi_{\varphi}(x)$$

$$\forall x \in \Omega \chi_{\theta \vee \varphi}(x) = \chi_{\theta}(x) + \chi_{\varphi}(x) - \chi_{\theta}(x) \times \chi_{\varphi}(x)$$

2.3.3.3 Likelihood Semantics

Likelihood semantics was first proposed as part of the TEE (Threshold, Error, Equivalence) model by Hisdal [44] and, as point out by Dubois and Prade [25], is closely related to random set semantics. Hisdal [44] suggests that membership functions can be derived from so-called yes-no experiments where a population of individuals are asked whether or not a particular label expression θ can be used to describe a certain value x . The membership of x in the extension of θ is then taken to be defined as follows:

$$\chi_{\theta}(x) = P(\theta|x)$$

where the above probability corresponds to the likelihood that a randomly chosen individual will respond with a yes to the question ‘is this value θ ?’ given that the value is x .

It is indeed not difficult to see connections between Hisdal’s yes-no experiments and the voting model described above. For instance, we might assume that an individual (voter) will repond yes that θ can describe x if and only if x is an element of that individual’s extension of θ generated as part of a voting experiment of the type discussed in the previous section. In this case

$$P(\theta|x) = P(\{v \in V : x \in R_{\theta}(v)\})$$

Given this relationship and indeed the fundamental properties of probability theory it is clear that fuzzy memberships based on likelihoods will not, in general, be truth-functional nor even functional.

Summary

In this chapter we have given an overview of the theory of fuzzy sets as proposed by Zadeh [111]. This theory identifies a formal truth-functional calculus for membership functions where the membership value of a compound fuzzy set is obtain by applying truth functions for the various connectives to the component membership functions. Truth-functionality was then compared with a weaker form of functionality in terms of the Boolean properties that can be satisfied by the resulting calculus.

The review of fuzzy set theory focused particularly on possible operational semantics for fuzzy membership functions and the consistency of each proposed semantics with Zadeh's truth-functional calculus was investigated. Indeed it is this semantic based analysis that highlights the real problem with truth-functional fuzzy set theory. For while a notion of membership function based on any of the interpretation discussed in this chapter may indeed prove to be a useful tool for modelling vague concepts none provide any convincing justification that the calculus for combining such membership function should be truth-functional. In other words, the real criticism of fuzzy logic is not that it fails to satisfy any particular Boolean property (as suggested by Elkan [30]) but rather that it has no operational semantics which is consistent with its truth-functional calculus. As such it fails to satisfy Walley's [103] interpretability principle for uncertainty measures.

There are two main responses to this criticism of fuzzy set theory. The first is that while no totally convincing semantics has been identified that justifies the truth-functionality assumption this does not mean that such a semantics will not be identified in the future. This is undeniably true but it will certainly necessitate that more attention is paid by the research community to the issue of operational semantics for membership functions than is currently the case. The second response is that the lack of a semantics does not matter since we can take membership values as primitives in the same way as crisp membership values are primitives. However, this position would seem somewhat hard to justify. For instance, unlike crisp sets there are no physical realisations of fuzzy sets. That is while there are crisp sets of objects that occur in the physical world, fuzzy sets, if they occur at all, can only really occur as subjective constructs within a certain linguistic context. For physical realisations of crisp sets membership functions are objective measurements of reality and it seems likely that it is from this connection to the physical world that many of our intuitions regarding the calculus for crisp memberships are derived. For fuzzy sets, even assuming truth-functionality based on the t-norm, t-conorm and negation axioms, there is

no consensus regarding the definition of truth functions for membership values. Indeed, as we have seen, there does not seem to be any intuitive justification for the truth-functionality assumption itself.

If membership functions for the extensions of vague concepts exist, or indeed even if they are ‘convenient fictions’ for modelling vagueness as suggested by Dubois and Prade [24], how can we evaluate them? Certainly we would not expect individuals to be able to reliably estimate their own fuzzy membership function for a vague concept through some process of introspection. Even if membership functions are in some way represented in an individual’s head, as suggested by Hajek [42], there would be no reason to suppose that they would have access to them (as is admitted by Hajek). In this case we require some behavioural mechanism according to which we can elicit membership function from individuals, this in itself requiring a lower-level understanding of membership functions and the way they are used.

Notes

- 1 One is tempted to say that either 'x is θ ' or that 'x is a θ ' should be a declarative statement but this would perhaps prejudge any discussion on the allocation of truth values to fuzzy statements
- 2 alternatively full compositionality [22] or strong functionality [67]
- 3 Weak Functionality in Lawry [67]
- 4 In many cases the following generalization of the triangle inequality is also required as property: $\forall x, y, z \in \Omega S(x, z) \geq f_{\wedge}(S(x, y), S(y, z))$ for some t-norm f_{\wedge} .