CHAPTER 3

EVIDENCE SETS: CONTEXTUAL CATEGORIES[16]

1. Cognitive Categorization

> "Most of our words and concepts designate categories. [...] Categorization is not a matter to be taken lightly. There is nothing more basic than categorization to our thought, perception, action, and speech. Every time we see something as a *kind* of thing, for example, a tree, we are categorizing. [...] An understanding of how we categorize is central to any understanding of how we think and how we function, and therefore central to an understanding of what makes us human". [Lakoff, 1987, pages xiii, 5, and 6]

Categories are bundles of concepts somehow associated in some context. Cognitive agents survive in a particular environment by categorizing their perceptions, feelings, thoughts, and language. The evolutionary value of categorization skills is related to the ability cognitive agents have to discriminate and group relevant events in their environments which may demand reactions necessary for their survival. If organisms can map a potentially infinite number of events in their environments to a relatively small number of categories of events demanding a particular reaction, and if this mapping allows them to respond effectively to relevant aspects of their environment, then only a finite amount of memory is necessary for an organism to respond to a potentially infinitely complex environment. In other words, only through effective categorization can knowledge exist in complicated environments.

Thus, knowledge is equated with the survival of organisms capable of using memories of categorization processes to choose suitable actions in different environmental contexts. It is not the purpose here to dwell into the interesting issues of evolutionary epistemology [Campbell, 1974; Lorenz, 1971]; I simply want to start this discussion by positioning categorization as a very important aspect of the survival of memory empowered organisms. Understanding categorization as an evolutionary (control) relationship between a memory empowered organism and its environment, implies the understanding of knowledge not as a completely observer independent mapping of real world categories into an organism's memory, but rather as the organism's, embodied, thus subjective, own construction of relevant – to its survival – distinctions in its environment. This is the basis for the evolutionary constructivist position discussed in chapter 2.

Since effective categorization of a potentially infinitely complex environment allows an organism to survive with limited amounts of memory, we can also see a connection between uncertainty and categorization. George Klir [1991] has argued that the utilization of uncertainty is an important tool to tackle complexity. If the embodiment of an organism allows it to recognize (construct) relevant events in its environment, but if all the recognizable events are still too complex to grasp by a limited memory system,

[16] The material in this chapter has been previously published in Rocha [1994a, 1995d, 1996b, 1997a, 1997b, 1997c], Rocha, Kreinovich and Kearfott [1996] and Henry and Rocha [1996].

the establishment of one-to-many relations between tokens of these events and the events themselves, might be advantageous for its survival. In other words, the introduction of uncertainty may be a necessity for systems with a limited amount of memory, in order to maintain relevant information about their environment. Thus, it is considered important for models of human categories to capture all recognized forms of uncertainty.

Lakoff [1987] has stressed the relevance of the idea of categories as subjective constructions of any beings doing the categorizing, and how it is at odds with the traditional objectivist scientific paradigms. In the following, I will address the historical relation between set theory and our understanding of categories; in particular, I will discuss what kind of extensions we need to impose on fuzzy sets so that they may become better tools in the modeling of subjective, uncertain, cognitive categories.

1.1 Models of Cognitive Categorization

It is important to separate the idea of a model of cognitive categorization and a model of a category. Though obviously dependent on one another, categories are included in more general models of cognitive categorization and knowledge representation. Agreeing on what the structure of a category might be, is far from agreeing on what the structure and workings of cognitive categorization models should be. It is also a simpler problem. Lakoff [1987], for instance, proposes a theory of knowledge organization based on structures called *idealized cognitive models (ICM)*, which contain categories as their substructures or by-products. Other similar models of knowledge organization exist: schema theory [Rumelhart, 1975], frames with defaults [Minsky, 1975], frame semantics [Fillmore, 1982], Dynamic Type Hierarchies [Way, 1991], etc. These models of knowledge organization vary in some ways; Lakoff's ICM's possess in addition to a propositional structure, a subject-dependent physiological structure, and metaphoric and metonymic mappings; in contrast to, for instance, Minsky’s purely propositional frames with defaults.

Though, undoubtedly, the specific model of knowledge organization selected will dictate some of the properties of categories, the particular structure chosen to represent categories in such models does not have to offer an explanation for knowledge organization. All that is asked of a good category representation, is that it may allow the larger imbedding model of knowledge representation to function. For instance, if we use mathematical sets to represent categories, our models of knowledge representation may use set theory connectives and/or they may use more complicated sets of mappings or even introduce connectionist machines to produce the sets [Clark, 1993]. Thus, evaluating sets as prospective representations of categories should be done by analyzing the kinds of limitations they <u>necessarily</u> impose on <u>any</u> kind of model, and not simply models circumscribed to basic set-theoretic operations.

1.2 The Classical View

The classical theory of categorization defines categories as containers of elements with common properties. Naturally, the classic, crisp, set structure was ideal to represent such containers: an element of a universe of observation can be either inside or outside a certain category, if it has or has not, respectively, the defining properties of the category in question. Further, all elements have equal standing in the category: there are no preferred representatives of a category – all or nothing membership.

One other characteristic of the classical view of categorization has to do with an observer independent epistemology: realism or objectivism. Cognitive categories were thought to represent objective distinctions in the real world; say, divisions between colors, between sounds, were all assumed to be characteristics of the real world independent from any beings doing the categorizing. Frequently, this objectivism is linked to the way classical categories are constructed on all-or-nothing sets of objects: "if categories are defined only by properties inherent in the members, then categories should be independent of the peculiarities of any beings doing the categorizing" [Lakoff, 1987, page 7]. I do not subscribe to this point of view; we can use classical categories both in realist or constructivist epistemologies. Even with classical, all-or-nothing, categories, the properties are never considered inherent in the members, there is always something defining the necessary list of properties: the external observer/constructor [Medina-Martins and Rocha, 1992]. The question is who or what is to establish the shared properties of a particular category. A model, where these shared properties are regarded as observer dependent, that is, established in reference to the particular physiology and cognition of the agent doing the categorizing, is built under a constructivist epistemology. If on the other hand, these properties are considered to be the one and ultimate truth of the real world, then the aim is the definition of an objectivist model of reality.

Most modern theories of categorization will include classical categories as a special case of a more complex scheme, which does not imply that some categories are objective and others are subjective. Thus, classical categories have to do with an all-or-nothing description of sets, based on a list of shared properties defined in some model. This external model is indeed built within an objectivist epistemology in the classical approach, but these two aspects of the classical theory of categorization are not necessarily dependent. The chosen structure of categories and the chosen model of knowledge representation/manipulation, which can be realist or constructivist, may be independent concerns when modeling cognitive categorization.

1.3 Prototype Theory and Fuzzy Sets

Rosch [1975, 1978] proposed a theory of category prototypes in which, basically, some elements are considered better representatives of a category than others. It was also shown that most categories cannot be defined by a mere listing of properties shared by all elements. Some approaches define this degree of representativeness as the distance to a salient example element of the category: a *prototype* [Medin and Schaffer, 1978]. More recently, prototypes have been accepted as abstract entities, and not necessarily a real element of the category [Smith and Medin, 1981]. An example would be the categorization of eggs by Lorenz'[1981] geese, who seem to use an abstract prototype element based on such attributes as color, speckled pattern, shape, and size. It is easy to fool a goose with a wooden egg if the abstract characteristics of the prototype are emphasized.

Naturally, fuzzy sets became candidates for the simulation of prototype categories on two counts: (i) membership degrees could represent the degree of prototypicality of a concept regarding a particular category; (ii) a category could also be defined as the degree to which its elements observe a number of properties, in particular, these properties may represent relevant characteristics of the prototype. These two points are distinct. The first makes no claim whatsoever on the mechanisms of creation and manipulation of categories. It may be challenged, as I will do in the sequel, on the grounds that due to its simplicity, models using it must be extremely complicated. Nonetheless, it does offer the minimum requirement a category must observe: a group (set) of elements with varying degrees of representativeness of the category itself.

Now, the second point goes beyond the definition of a category and enters the domain of modeling the creation of categories. As in the classic case, categories are seen as groups of elements observing a list of properties, the only difference is that elements are allowed to observe these properties to a degree. However, the so called radial categories [Lakoff, 1987] cannot be formed by a listing of properties shared by all its elements, even if to a degree. They refer to categories possessing a central subcategory core, defined by some coherent (to a model or context) listing of properties, plus some other elements which must be learned one by one once different contexts are introduced, but which are unpredictable from the core's context and its listing of shared properties[17]. Thus, the second interpretation of fuzzy sets as categories leads fuzzy logic to a corner which renders it uninteresting to the modeling of cognitive categorization. Notice that Rosch herself made a distinction between the notion of category prototypes and the notion of knowledge representation:

> "Prototypes do not constitute any particular processing model for categories [...]. What the facts about prototypicality do contribute to processing notions is a constraint ─ process models should not be inconsistent with the known facts about prototypes. [...] As with processing models, the facts about prototypes can only constrain, but do not determine, models of representation." [Rosch, 1978, pg. 40]

1.4 Dynamic Categories

As Hampton [1992] and Clark [1993] discuss, the important question to ask at this point is “where do the distance degrees come from?” Barsalou [1987] has shown how the prototypical judgments of categories are very unstable across contexts. He proposes that these judgements, and therefore the structure of categories, are constructed “on the hoof” from contextual subsets of information stored in long-term memory. The conclusion is that such a wide variety of context-adapting categories cannot be stored in our brains, they are instead dynamic categories which are rarely, if ever, constructed twice by the same cognitive system. Categories have indeed Rosch’s graded prototypicality structure, but they are not stored as such, merely constructed “on the hoof” from some other form of information storage system.

> “Invariant representations of categories do not exist in human cognitive systems. Instead, invariant representations of categories are analytic fictions created by those who study them.” [Barsalou, 1987, page 114]

As Clark [1993] points out, the reason for this is that since the evidence for graded categories is so strong, even in ad hoc categories such as “things that could fall on your head” or viewpoint-related categories, “it seems implausible to suppose that the gradations are built into some preexisting conceptual unit or prototype that has been simply extracted whole out of long-term memory.” [Ibid, page 93] Thus, we should take the graded prototypical categories as representations of these highly transient, context-dependent knowledge arrangements, and not of models of information storage in the brain. In the following, the extensions of fuzzy sets proposed to model cognitive categories should be understood as such.

As for the modeling of cognitive categorization itself, an attempt to model certain aspects of it is developed with an extended theory of approximate reasoning, which is used on a computational system of database retrieval developed in chapter 5. In section 6, this extended theory of approximate reasoning is developed.

[17] An example of a radial category [after Lakoff, 1987] is the category of mother. A listing of core properties to be considered a member of this category, coherent in the context of birth, would be, for instance: woman who gives birth, raises, nurtures, educates a child. However, members of the category of mother exist which do no obey such listing: adoptive mother, surrogate mother, genetic mother, etc. These members do not obey the list, or obey it only to a very small degree; however, though not prototypes, they are elements of the category mother. They are also not random elements, but are unpredictable until a different context is introduced.

1.5 Fuzzy Objectivism

With fuzzy sets and approximate reasoning Zadeh [1965, 1971] substitutes a classic logic of truth by a logic of degrees of truth: instead of having members of classes/categories which belong or not belong to it, we have members that possibly belong to a category to a certain degree. Lakoff [1987] believes that the utilization of degrees of truth adds nothing to the main shortcoming of classical categories, as they are usually thought of as objective graded degrees that exist in the real world; objectivism is merely replaced by fuzzy objectivism. Now, even if Zadeh's initial formulation of fuzzy sets may have been indeed a realist one, nothing prevents us from using fuzzy sets as representations of categories within a constructivist epistemology. Categories defined by fuzzy sets may represent degrees of prototypicality which may vary according to contexts introduced in imbedding models of categorization processes. In particular, a model may take into account levels of physiological subjectivity as desired by Lakoff [1987]. A computational example of such a model has been developed by Medina-Martins and Rocha [1992; Medina-Martins, Rocha, et al, 1994; Medina-Martins, 1995].

Since fuzzy sets, at least to a degree, can be included in realist or constructivist frameworks, its dismissal as good models of cognitive categories has to be made on different grounds. In the following I will maintain that fuzzy sets are unsatisfactory because they (i) lead to very complicated models, (ii) do not capture all forms of uncertainty necessary to model mental behavior, and (iii) leave all the considerations of a logic of subjective belief to the larger imbedding model, which makes them poor tools in evolutionary constructivist approaches. A formal extension based on evidence theory is proposed next.

2. Mathematical Background

Let X denote a nonempty *universal set* under consideration. Let $\mathscr{P}(X)$ denote the power set of X. An element of X represents a possible value for a variable x. X can be *countable* or *uncountable*. The term continuous domain is often used to refer to the latter case. However, since continuity is a notion applicable to functions not sets, I will use the term "uncountable set" to refer to those sets with uncountably many elements. An uncountable set is by definition an infinite set, a countable set can be both finite or infinite.

2.1 Measures

Let μ be an extended real valued measure on a σ-algebra[18] $\mathscr{E}$ of X: μ: $\mathscr{E} \rightarrow [0, \infty]$. μ is *countably additive* for any numerable disjoint sequence $\{A_n\}$ of sets in $\mathscr{E}$:

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) \quad \textbf{(1)}$$

[18] A σ-algebra is a class of X that contains X and is closed under the formation of complements and countable unions.

also, $\exists\ A \subseteq X$ such that $\mu\ (A) < \infty$ and $\mu(\varnothing)=0$, From the additivity requirement (1) the following *monotonicity condition* is obtained: $\mu(A) \leq \mu(B)$, when $A \subseteq B$. The monotonicity further implies that: $\max \mu(A) = \mu(X)$, and $\min \mu(A) = \mu(\varnothing) = 0$, where $A \subseteq X$.

Since $\mathscr{E}$ is closed to the intersection of its elements, and since μ is additive and monotone, it is also *countably subadditive* for any sequence $\{A_i\}$ of sets in $\mathscr{E}$ such that $A = \bigcup_{i=1}^{\infty} A_i$:

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mu(A_i) \qquad \textbf{(2)}$$

A *nonadditive measure* (or *fuzzy measure*) is defined by the same boundary conditions and condition of monotonicity, but the additivity requirement is relaxed to one of continuity. See Wang and Klir [1992] for details unnecessary for the purposes of this dissertation.

2.2. Dempster-Shafer Theory of Evidence

2.2.1 Basic Probability Assignment

Evidence theory, or *Dempster-Shafer Theory* (DST) [Shafer, 1976], may be defined in terms of a set function m:

$$m: \mathscr{P}(X) \rightarrow [0, 1]$$

referred to as a *basic probability assignment*, such that $m(\varnothing) = 0$ and

$$\sum_{A \subseteq X} m(A) = 1 \qquad \textbf{(3)}$$

The value $m(A)$ denotes the proportion of all available evidence which supports the claim that $A \in \mathscr{P}(X)$ approximately represents the actual value of our variable x. $m(A)$ qualifies A alone, it does not imply any additional claims regarding other subsets of X, including subsets of A or the complement of A.

2.2.2 Belief and Plausibility

DST is based on a pair of nonadditive measures: *belief* (Bel) and *plausibility* (Pl). These measures observe axioms of *superadditivity* and *subadditivity* respectively [for more details see Wang and Klir, 1992]. Given a basic probability assignment m, Bel and Pl are determined for all $A \in \mathscr{P}(X)$ by the equations:

$$\mathrm{Bel}(A) = \sum_{B \subseteq A} m(B) \qquad \textbf{(4)}$$

$$\mathrm{Pl}(A) = \sum_{B \cap A \neq \varnothing} m(B) \qquad \textbf{(5)}$$

(3), (4), and (5) imply that belief and plausibility are dual measures related by:

$$\mathrm{Pl}(A) = 1 - \mathrm{Bel}(A^c) \quad \textbf{(6)}$$

for all $A \in \mathcal{P}(X)$, where A^c represents the complement of A in X. It is also true that $\mathrm{Bel}(A) \leq \mathrm{Pl}(A)$ for all $A \in \mathcal{P}(X)$. Notice that [Shafer, 1976, page 38] , "$m(A)$ measures the belief one commits exactly to A, not the total belief that one commits to A." Bel(A), the total belief committed to A, is instead given by the sum of all the values of m for all subsets of A (4).

2.2.3 Focal Elements and Bodies of Evidence

Any set $A \in \mathcal{P}(X)$ with $m(A) > 0$ is called a *focal element*. A *body of evidence* is defined by the pair $(\mathfrak{F}, m)$, where $\mathfrak{F}$ represents the set of all focal elements in X, and m the associated basic probability assignment. $\mathfrak{F}$ is assumed to be finite, that is, there is a finite number of focal elements in a body of evidence, even if the universal set X is infinite. This is an important consideration which is utilized throughout the dissertation. Kramosil [1995] has developed a more complete extension of DST to infinite domains which considers cases where the set of all focal elements in X, $\mathfrak{F}$, is infinite. Since the applications I have in mind require only a finite number of focal elements, I will not consider extensions of DST such as those of Kramosil.

The set of all bodies of evidence is denoted by $\mathfrak{B}$. *Total ignorance* is expressed in DST by $m(X) = 1$ and $m(A) = 0$ for all $A \neq X$. *Full Certainty* is expressed by $m(\{x\}) = 1$ for one particular element of X, and $m(A) = 0$ for all $A \neq \{x\}$.

2.2.4 Dempster's rule of combination

In the context of evidence theory, the universal set X is referred to as the *frame of discernment*. Given two pairs of dual belief-plausibility measures, Bel_1-Pl_1 Bel_2-Pl_2, over the same of frame of discernment X, but based on different bodies of evidence $(\mathfrak{F}, m)_1$, $(\mathfrak{F}, m)_2$, the resulting, combined, body of evidence, $(\mathfrak{F}, m)_{1,2}$, is defined by the following basic probability assignment:

$$m_{1,2}(C) = \frac{\sum_{A_i \cap B_j = A} m_1(A_i)\, m_2(B_j)}{1 - \sum_{A_i \cap B_j = \emptyset} m_1(A_i)\, m_2(B_j)} \quad \textbf{(7)}$$

where $\mathfrak{F}_{1,2}$ is the set of all non-empty subsets C of X resulting from the intersection of each focal element A_i of $\mathfrak{F}_1$ with each focal element B_j of $\mathfrak{F}_2$. (7) is referred to as the *Dempster's rule of combination*.

2.2.5 Joint Bodies of Evidence

We further need to understand properties of bodies of evidence defined on the Cartesian product of two sets: $Z = X \times Y$. That is, bodies of evidence defined by *joint basic probability assignment* functions: m: $\mathcal{P}(X \times Y) \rightarrow [0, 1]$, obeying equation (3) above, where X and Y denote sets defining the domain of two distinct variables. The focal elements are binary relations, R, defined on Z. The *projections* of R on X and Y are given, respectively, by the sets:

$$R_X = \{x \in X \mid (x,y) \in R \text{ for some } y \in Y\}$$

$$R_Y = \{y \in Y \mid (x,y) \in R \text{ for some } x \in X\}.$$

From these we can calculate the *marginal basic probability assignments*, m_X and m_Y, from the joint basic probability assignment m:

$$m_X(A) = \sum_{R|A=R_X} m(R) \quad \forall A \subseteq X$$

$$m_Y(B) = \sum_{R|B=R_Y} m(R) \quad \forall B \subseteq Y$$

Two bodies of evidence $(\mathfrak{F}_X, m_X)$ and $(\mathfrak{F}_Y, m_Y)$ are *noninteractive* if and only if for all $A \in \mathfrak{F}_X$ and for all $B \in \mathfrak{F}_Y$:

$$m(A \times B) = m_X(A) \cdot m_Y(B) \tag{8}$$

and $m(R) = 0$ for all $R \neq A \times B$. Noninteractive joint bodies of evidence can be completely recuperated from their projected marginal bodies of evidence: $R = R_X \times R_Y$. In general we have instead: $R \subseteq R_x \times R_Y$[19].

2.2.6 Inclusion

Following Dubois and Prade [1986, 1987], we can also consider the idea of *inclusion* in DST. Given two bodies of evidence $(\mathfrak{F}_1, m_1)$ and $(\mathfrak{F}_2, m_2)$ on X, $(\mathfrak{F}_1, m_1) \subseteq (\mathfrak{F}_2, m_2)$ if and only if:

(i) $\forall A \in \mathfrak{F}_1$, $\exists B \in \mathfrak{F}_2$, such that $A \subseteq B$,

(ii) $\forall B \in \mathfrak{F}_2$, $\exists A \in \mathfrak{F}_1$, such that $A \subseteq B$,

(iii) $\exists w: \mathscr{P}(X \times X) \rightarrow [0, 1]$, with $w(A, B) = 0 \; \forall A \notin \mathfrak{F}_1$, or $B \notin \mathfrak{F}_2$,
such that $\forall A \subseteq X,\; m_1(A) = \sum_{B|A\subseteq B} w(A,B)$, $\forall B \subseteq X,\; m_2(B) = \sum_{A|A\subseteq B} w(A,B)$

2.3 Fuzzy Sets and Interval Valued Fuzzy Sets

A crisp set entails no uncertainty in its membership assessment: if an element x of X is a member of a set $A \subseteq X$, then it will not be a member of its complement $A^c \subseteq X$. A fuzzy set introduces fuzziness as

[19] To visualize this, consider an "L" shaped focal element of $X \times Y$, its projections R_X and R_Y will be the same as those of a focal element given by a rectangle with the same width and length as the L-shaped one. The rectangle clearly contains the "L"-shaped focal element.

the above law of contradiction is violated: x can both be a member (to a degree) of A and A^c. A fuzzy set[20] A is defined by a membership function $A\colon\ X \rightarrow [0,1]$.

Fuzzy sets can be extended to *interval valued fuzzy sets* (IVFS), a case of *probabilistic sets* [Hirota, 1981] and *type 2 fuzzy sets* [Zadeh, 1975]. A probabilistic set A assigns to each element, x of X, a probability distribution defined on [0, 1] describing its membership: $A\colon\ X \rightarrow ([0,1], \boldsymbol{P})$. $\boldsymbol{P}$ represents the set of all probability distributions on [0,1]. A type 2 fuzzy set assigns to each element x of X , a fuzzy set, defined on [0, 1]: □: $X \rightarrow$ $\boldsymbol{F}$([0,1]). $\boldsymbol{F}$([0,1]) is the set of all possible fuzzy sets defined on [0,1]. However, we do not need a probabilistic/possibilistic representation to define an IVFS; all we need is to assign an interval of [0, 1] to each element x of X : □: $X \rightarrow \mathcal{I}$([0,1]), where $\mathcal{I}$ represents the set of intervals in [0, 1]. The greatest advantage of IVFS over probability/possibilistic sets is their comparative simplicity: probability or possibility distributions need an extensive and precise amount of information that is usually not obtainable in relevant applications. Conversely, an IVFS does not require much more information than a fuzzy set, simply an upper and a lower limit to its interval of membership. Moreover, an IVFS offers a kind of uncertainty fuzzy sets cannot capture and which probabilistic and possibilistic sets initially aimed at [Hirota, 1981]. Basically, it is not always possible to unequivocally specify a single membership value for some linguistic category we wish to model [Gorzałczany, 1987], as fuzzy sets demand.

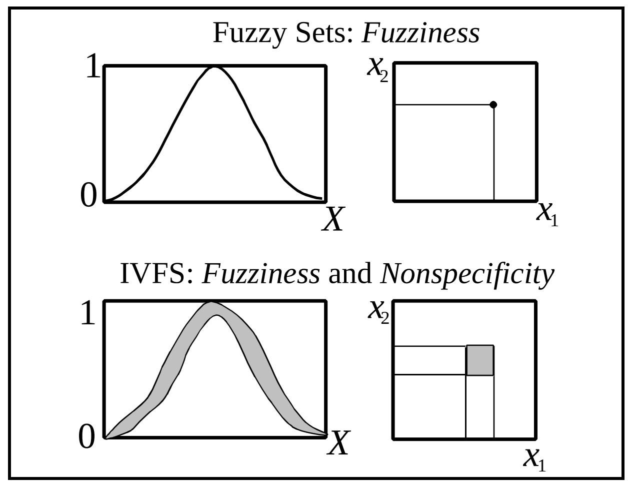

Figure 1: Fuzzy Sets, IVFS, and Uncertainty. The right column uses Kosko's hypercube representation of a fuzzy set with 2 elements.

IVFS offer, in addition to fuzziness, a nonspecific description of membership in a set; and they do so with very little information requirements. An IVFS A, <u>for each x in X</u>, captures two forms of uncertainty (see below): *fuzziness* (as in the case of normal fuzzy set) and *nonspecificity*. The membership degrees of standard fuzzy sets are absolutely specific. When we create a fuzzy set we have perfect knowledge of the degree to which a certain element x of X belongs to A. In contrast, when we create an IVFS we have nonspecific knowledge of the degree of membership; hence the utilization of an interval to describe the membership of x in A[21]

2.4 Uncertainty

The notion of uncertainty, is very relevant to any discussion of the modeling of linguistic/mental abilities. From Zadeh's [1971, 1975, 1978] approximate reasoning to probabilistic and even evidential

[20] This is the definition of a standard fuzzy set. Other types of fuzzy sets exist, some of which are introduced throughout this chapter. In the following, unless otherwise noticed, the name fuzzy set refers to this definition of standard fuzzy set.

[21] Notice that both crisp and fuzzy sets capture nonspecificity in the cardinality of their elements. The inclusion of nonspecificity discussed here is in the formalization of their membership degrees, which in this sense, represents a second-order nonspecificity.

reasoning [Schum, 1994], uncertainty is more and more recognized as a very important issue in cognitive science and artificial intelligence with respect to the problems of knowledge representation and the modeling of reasoning abilities [Shafer and Pearl, 1990]. Engineers of knowledge based systems can no longer be solely concerned with issues of linguistic or cognitive *representation*, they must describe "*reasoning*" procedures which enable an artificial system to answer queries. In many artificial intelligence systems, the choice of the next step in a reasoning procedure is based upon the measurement of the system's current uncertainty state [Nakamura and Iwai, 1982; Medina-Martins and Rocha, 1992, Medina-Martins et al, 1993, 1994; see chapter 5]. Thus, we need to collect an array of effective uncertainty measures in order to improve current models of knowledge representation.

George Klir [1993; Klir and Yuan, 1995] classifies uncertainty into two main forms: *ambiguity* and *fuzziness*. Ambiguity is further divided into the categories of *nonspecificity* and *conflict*. Webster [Random House Webster's Dictionary, 1991] defines ambiguity as: *1. doubtfulness or uncertainty of meaning or intention: to speak with ambiguity. 2. the condition of admitting more than one meaning.* Mathematically ambiguity is identified with the existence of one-to-many relations, that is, when several alternatives exist for the same question or proposition. DST provides an ideal framework for the study of ambiguity, as it enlarges the scope of traditional probability theory, and it can be interpreted in terms of the possible-world semantics of classical modal logics [Resconi, et al, 1993]. As measures of ambiguity, we are looking for functions of the form: f: $\mathscr{B} \rightarrow [0, \infty)$, where $\mathscr{B}$ is the set of bodies of evidence defined on X.

Notice that the measures of ambiguity f, or more generally measures of uncertainty based information [Klir, 1993], are not classical measures as defined in section 2.1. The former are functions defined on bodies of evidence, while the latter are set functions with the axioms of section 2.1. Indeed, even though sharing the same term 'measure', they refer to quite distinct mathematical concepts. Classical measures are associated with the definition of metric spaces and topologies, while uncertainty measures are defined to capture amounts of information in uncertain situations. In this dissertation, the uncertainty framework chosen to deal with ambiguity is the Dempster-Shafer theory of evidence, thus, uncertainty measures are defined on bodies of evidence. In other words, measures of uncertainty require first and foremost a mathematical framework for uncertainty, while measure theory is based on set theory alone. Furthermore, as it will be discussed in the following sections, measures of uncertainty are often required to follow an axiom of additivity quite distinct from the additivity requirement of a classical measure. The two should not to be confused since they refer to different mathematical concepts that unfortunately share the same names.

2.4.1 Conflict

Conflict is identified with disagreement between several alternatives. The word disagreement implies the existence of some distinctive criteria between the several alternatives. When it is possible to distinguish between the several alternatives of some event or proposition, we have conflict amongst alternatives. In probability theory, conflict is measured with the Shannon measure of entropy. In evidence theory, the probabilistic entropy is efficiently extended to a measure of uncertainty named *Strife* [Vejnarová and Klir, 1993]:

$$S(m) = -\sum_{A \in \mathscr{F}} m(A) \log_2 \sum_{B \in \mathscr{F}} m(B) \frac{|A \cap B|}{|A|} \qquad \textbf{(9)}$$

This measure was developed from another measure used to quantify conflict in DST named *discord* which was originally proposed by Klir and Ramer [1990]. Discord, instead of using the degree of subsethood of set A in B ($|A \cap B| / |A|$), uses the degree of subsethood of set B in A ($|A \cap B| / |B|$).Arguments for introducing strife are discussed by Klir and Parviz [1992]. Klir and Yuan [1993] have observed that the distinction between strife and discord reflects the distinction between disjunctive and conjunctive set valued statements [Klir and Wierman, 1997]. In the following only the measure of strife is used, since all developments proposed for the measure of strife can be trivially extended to discord.

2.4.2 Nonspecificity

Nonspecificity, is identified with unspecified distinctions between several alternatives, that is, when we possess several alternatives which are equally possible, or probable. "[Nonspecificity] is connected with sizes (cardinalities) of relevant sets of alternatives." [Klir, 1993, page 276] In the formal domain, this is expressed by the weighted average of the Hartley measure of information:

$$N(m) = \sum_{A \in \mathcal{F}} m(A) \log_2 |A| \tag{10}$$

2.4.3 Fuzziness

Fuzziness is usually identified with lack of sharp distinctions; other synonyms of the word *fuzzy* include: *blurred, indistinct, unclear, vague, ill-defined, out of focus, not clear, indefinite; shadowy, dim, obscure; misty, hazy, murky, foggy; confused* [Random House Webster's Dictionary, 1991]. Fuzzy sets are usually used to formalize this kind of uncertainty. The elements of a fuzzy set are in it included according to a membership degree between 0 and 1. In Fuzzy Logic terms, the truth value of a proposition, now a possibility value, ranges between 0 and 1. There are several ways of measuring fuzziness [Klir, 1993], but the most modern approaches define a measure of fuzziness as the lack of distinction between a set and its complement [Yager, 1979, 1980]. "Indeed, it is precisely the lack of distinction between sets and their complements that distinguishes fuzzy sets from crisp sets. The less a set differs from its complement, the fuzzier it is." [Klir, 1993, page 298] In other words, the more something is and is not, <u>at the same time</u>, the fuzzier it is. As truth is substituted for possibility, we find that what is possible to a degree, is also not possible to the inverse degree— in the limit, when something is possible and not possible to the same degree (½) we have paradox[22]. A measure of fuzziness usually defined within this interpretation of fuzziness is given by:

$$F(A) = \sum_{x \in X} \left[1 - |2\mu_A(x) - 1| \right] \tag{11}$$

which sums the lack of distinction between the membership of each element x of the universal set X, in a fuzzy set A and in its complement.

[22]Possibility understood as the truth-value representation of fuzzy logic propositions, not to normalized possibility distributions in the current interpretations of possibility theory [de Cooman et al, 1995].

3. Sets and Cognitive Categorization

3.1 Fuzzy Sets and the *Prototype Combination Problem*

Whenever fuzzy set models of cognitive categories have been proposed, a model of cognitive categorization or human reasoning has also been included in the package. Zadeh [1975] proposed a theory of approximate reasoning based of fuzzy predicate logic. Gorzałczany [1987] proposed a method of inference in approximate reasoning based on interval-valued fuzzy sets. Turksen [1986] presented a method of concept combination based on the idea that fuzzy sets, when combined, should introduce a second degree level of uncertainty. Bo Yuan et al [1994] also investigate an interval valued fuzzy set approach to approximate reasoning based on normal forms. Atanassov [1986; Atanassov and Gargov, 1989] introduced the concept of *intuitionistic fuzzy sets* and *intuitionistic interval valued fuzzy sets* together with a whole set of operators [Atanassov, 1994] leading to yet another form of approximate reasoning. These are some of the available models of fuzzy reasoning based on fuzzy categories.

As previously discussed, fuzzy sets are actually fairly accurate representations of categories simply because they are able to represent prototypicality (understood as degree of representativeness); how the prototype degrees are constructed is, on the other hand, a different matter. Fuzzy sets are simple representations of categories which need much more complicated models of approximate reasoning than those fuzzy predicate logic alone can provide in order to satisfactorily model cognitive categorization processes. Critics [Osherson and Smith, 1981; Smith and Osherson, 1984; Lakoff, 1987] have shown that the several fuzzy logic connectives (e.g. conjunction and disjunction) based on different conjugate pairs of t-norms and t-conorms[23], cannot conveniently account for the prototypicality of the elements of a complex category, which may depend only partially on the prototypicality of these elements in several of its constituent categories and may even be larger (or smaller) than in any of these. This is know as the *prototype combination problem*.

A complex category is assumed to be formed by the connection of several other categories. Approximate reasoning defines the sort of operations that can be used to instantiate this association. Smith and Osherson's [1984] results, showed that a <u>single</u> fuzzy connective cannot model the association of entire categories into more complex ones. Their analysis centered on the traditional fuzzy set connectives of (max-min) union and intersection. They observed that max-min rules cannot account for the membership degrees of elements of a complex category which may be lower than the minimum or higher than the maximum of their membership degrees in the constituent categories. Their analysis is very incomplete regarding the full-scope of fuzzy set connectives, since we can use other operators [see Dubois and Prade, 1985], to obtain any desired value of membership in the [0, 1] interval of membership. However, their basic criticism remains valid: even if we find an appropriate fuzzy set connective for a particular element, this connective will not yield an accurate value of membership for other elements of the same category. Hence, a model of cognitive categorization which uses fuzzy sets as categories will need several fuzzy set connectives to associate two categories into a more complex one (in the limit, one for each element). Such model will have to define the mechanisms which choose an appropriate connective for <u>each</u> element of a category. Therefore, a model of cognitive categorization based solely on fuzzy sets and their connectives will be very complicated and cumbersome. No single fuzzy set connective can account for the exceptions of different contexts, thus the

[23]Triangular norms (t-norms) and triangular conorms (t-conorms) are the general names given to the families of fuzzy intersections and unions respectively. Those t-norms that uniquely determine a dual t-conorm, and vice versa are referred to as conjugate pairs of t-norms and t-conorms.

necessity of a complex model which recognizes these several contexts before applying a particular connective to a particular element.

The prototype combination problem is not only a problem for fuzzy set models, but for all models of combination of prototype-based categories. Fodor [1981] insists that though it is true that prototype effects obviously occur in human cognitive processes, such structures cannot be fundamental for complex cognitive processes (high level associations): “there may, for example, be prototypical cities (London, Athens, Rome, New York); there may even be prototypical *American Cities* (New York, Chicago, Los Angeles); but there are surely not prototypical *American cities situated on the east coast just a little south of Tennessee*.”[Ibid, page 297] As Clark [1993] points out, the problem with Fodor’s point of view, and indeed the reason why fuzzy set combination of categories fails, is that “he assumes that prototype combination, if it is to occur, must consist in the linear addition of the properties of each contributing prototype.” [Ibid, page 107] Clark proposes the use of connectionist prototype extraction as an easy way out of this problem. In fact, a neural network trained to recognize certain prototype patterns, e.g. some representation of “tea” and “soft drink”, which is also able to represent a more complex category such as “ice tea”, “does not do so by simply combining properties of the two ‘constituent’ prototypes. Instead, the webs of knowledge structure associated with each ‘hot spot’ engage in a delicate process of mutual activation and inhibition.” [Ibid, page 107] In other words, complex categories are formed by nonlinear, emergent, prototype combination.

As Clark himself points out, however, this ability to nonlinearly combine prototypes in connectionist machines is a result of the pre-existence of a (loosely speaking) semantic metric which relates all knowledge stored in the network. It is not a proper metric since it may not follow the triangle inequality, but the kind of distance in which the shortest distance between two stored concepts may not be the straight line, often referred to as semi-metric. In any case, through the workings of the network with its inhibition and activation signals, new concepts can be learned which must somehow relate to the existing knowledge previously stored. Therefore, any new knowledge that a connectionist device gains, must be somehow related to previous knowledge. This dependence prevents the sort of open-ended conceptual combination that we require of higher cognitive processes.

This problem might be rephrased by saying that connectionist devices can only make nonlinear prototype combinations given a small number of contexts. We often use a network to classify, say, sounds, another one images, and so sorth. In their own contexts, each network combines prototypes into more complex ones, but they cannot escape their own contexts. I believe, with Clark, that connectionist machines are nonetheless very powerful, even given these constraints. The approach I am about to follow, is not proposed to be used instead of connectionist devices, but one that may offer a more high-level treatment of the contextual problem in prototype combination. In fact, in chapter 5, a computational model is presented that even though not using connectionist machines (distributed memory), uses networked relational databases that also possess semantic semi-metrics and which can approach this contextual problem.

3.2 Interval Valued Fuzzy Sets

As discussed in the previous section, fuzzy sets have extremely limited abilities to model the combination of prototypical categories. They can only work on very limited contexts, whose categories can be formed from the linear combination of constituent categories. The Introduction of a theory of approximate reasoning based on interval valued fuzzy sets [Gorzałczany, 1987; Türkşen, 1986] represents a step forward in the modeling of cognitive categorization, as it offers a second level of uncertainty, but it only slightly improves the contextual problem referred above. The membership degrees of IVFS are nonspecific (see section 2.3). This second dimension of uncertainty allows us to interpret the interval of membership of an element in a category as the membership degree of this element *according* to several different contexts, which we cannot a priori identify.

In particular, Turksen's concept combination mechanisms are based on the separation of the disjunctive and conjunctive normal forms of logic compositions in fuzzy logic. A disjunctive normal form (DNF) is formed with the disjunction of some of the four primary conjunctions, and the conjunctive normal form (CNF) is formed with the conjunction of some of the four primary disjunctions: $A \cap B$, $A \cap \overline{B}$, $\overline{A} \cap B$, $\overline{A} \cap \overline{B}$ and $A \cup B$, $A \cup \overline{B}$, $\overline{A} \cup B$, $\overline{A} \cup \overline{B}$. In two-valued logic the CNF and DNF of a logic composition are equivalent: CNF = DNF. Turksen observed that in fuzzy logic, for certain families of conjugate pairs of t-norms and t-conorms, we have instead DNF $\subseteq$ CNF for some of the fuzzy logic connectives. He then proposed that fuzzy logic compositions could be represented by IVFS's given by the interval [DNF, CNF] of the fuzzy set connective chosen [Turksen, 1986]. With IVFS based connectives, Turksen was able to deal more effectively with the shortcomings of a pure fuzzy set approach. In his model, two fuzzy sets are combined into an IVFS. The fuzzy and nonspecific degrees of membership of the elements in the category obtained, can be interpreted as inclusion in a category according to several possible, fuzzy degrees.

Turksen's model simplifies the pure fuzzy set approach since we will find more categories which can be combined into complex categories with a single connective used for all elements of the universal set, though it will not work for all radial categories. The IVFS approach provides a way to acknowledge the existence of contextual nonspecificity in complex category formation, thus producing a more accurate representation of different forms of uncertainty present in such processes. The problem is that categories demand membership values which more than nonspecific can be conflicting. That is, the contextual effects may need more than an interval of variance to be accurately represented. Also, even though IVFS use nonspecific membership, thus allowing a certain amount of contextual variance, the several contexts are not explicitly accounted for in the categorical representation. Section 4 proposes set structures which (i) capture all recognizable forms of uncertainty in their membership representation, (ii) point explicitly to the contexts responsible for a certain facet of their membership representation, and (iii) in so doing, introduce a formalization of belief.

3.3. Set complement and intuitionistic sets

Before I introduce such structures in section 4, a comment should be made regarding Atanassov's [1986, 1994; Atanassov and Gargov, 1989] *intuitionistic fuzzy sets* and *interval valued intuitionistic fuzzy sets*. A fuzzy set is defined by a degree of membership in [0,1]. As it was noticed in the discussion of uncertainty forms in section 3.2., fuzziness is identified with the conflict between inclusion and non-inclusion in a set. If an element *x* of *X* is included in set *A* to a degree *d*, then it is also not included in *A* to a (1-*d*) degree; in other words, it is included in the complement of *A*, to a (1-*d*) degree. An intuitionistic *A* set is instead defined by both the degree of membership and the degree of non-membership: $A = \left\{ \langle x, \mu_A(x), \nu_A(x) \rangle : x \in X,\ \mu_A, \nu_A : X \to [0,1] \right\}$, with the restriction: $\mu_A(x) + \nu_A(x) \leq 1$. This idea of intuitionistic set introduces an asymmetry between inclusion and non-inclusion which may be very relevant in the modeling of cognitive categories rich in all sorts of asymmetries. These sets have been successfully extended into IVFS with a whole set of relevant operators, thus Gorzałczany and Türkşen's mechanisms mentioned before could be endowed with this extra asymmetry. I do not pursue this avenue here, but it may prove to be an extension well worth pursuing.

4 Evidence Sets: Membership and Belief

An alternative way to represent an IVFS A is to consider that for every element x of X, there is a body of evidence $(\mathcal{F}^x, m^x)$ defined on the set of all intervals of [0,1], $\mathcal{I}[0, 1]$, with a single focal element given by the interval $I^x = [l^x_{inf}, l^x_{sup}] \subseteq [0,1]$. The basic probability assignment function m^x assumes the value **1** for this single focal element, representing our belief that the degree of membership of element x of X in A is (with all certainty) in the sub-interval I^x of [0, 1]. In other words, our *judgement* of the (nonspecific) degree of membership, I^x, of x in set A indicates that we fully believe it is correct. Notice that the universal set of the IVFS is X, but the universal set of the body of evidence is the unit interval[0, 1].

It is now clear that an IVFS is a very special case of a more general structure which I refer to as ***evidence set***. An *evidence set A* of X, is defined by a membership function of the form:

$$A(x): X \to \mathcal{B}[0, 1]$$

where, $\mathcal{B}[0, 1]$ is the set of all possible bodies of evidence $(\mathcal{F}^x, m^x)$ on $\mathcal{I}[0, 1]$. Such bodies of evidence are defined by a basic probability assignment m^x on $\mathcal{I}([0, 1])$, for every x in X (focal elements must be intervals). Notice that [0, 1] is an infinite, uncountable, set, while X can be countable or uncountable. Thus, evidence sets are set structures which provide interval degrees of membership, weighted by the probability constraint of DST. They are defined by two complementary dimensions: membership and belief. The first represents a fuzzy, nonspecific, degree of membership, and the second a subjective degree of belief on that membership, which introduces conflict of evidence as several, subjectively defined, competing membership intervals weighted by the basic probability constraint are created.

4.1 Consonant Evidence Sets

An interesting case occurs when we restrict $\mathcal{F}^x$ to consonant bodies of evidence, that is, to a nested structure of interval focal elements: $I^x_1 \subseteq I^x_2 \subseteq \cdots \subseteq I^x_n \subseteq [0,1]$. In this instance we obtain a sort of graded and nested structure of several IVFS (Figure 2), which leads to consonant belief measures: $Bel(I^x_1) \le Bel(I^x_2) \le \cdots \le Bel(I^x_n) \le 1$. Instead of using a single interval with maximum degree of belief, to formalize the nonspecificity of the degree of membership of element x of X in a set A, as is the case of IVFS, a consonant evidence set uses several nested intervals (three in the case of Figure 2) with different degrees of belief, stating our graded evidence claims regarding the membership of element x of X in A.

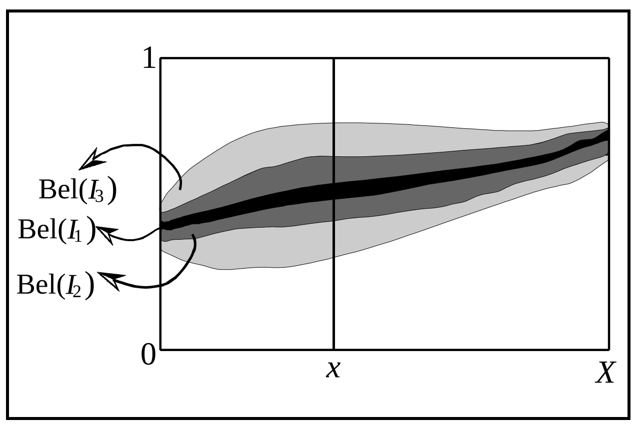

Figure 2: Consonant Evidence Set with 3 focal elements

4.2 Non-Consonant Evidence Sets

When $\mathfrak{F}^x$ is no longer restricted to consonant bodies of evidence, we obtain evidence sets that are a bit more "incoherent", that is, disjoint intervals of membership exist for the same membership degree in the evidence set. In other words, the evidence we possess leads to a conflicting characterization of the membership value of *x*. Figure 3 shows an example of a non-consonant evidence set.

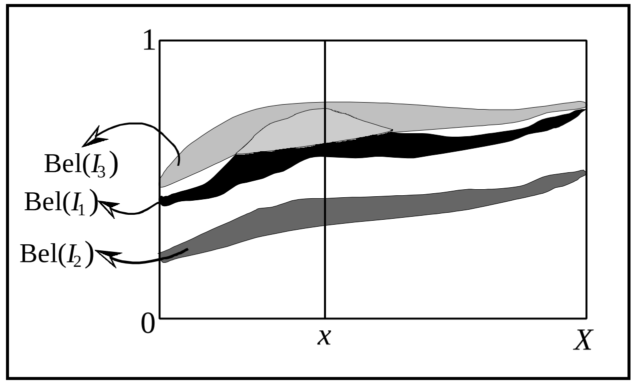

Figure 3: Non-Consonant Evidence Set with 3 focal elements

4.3 Complexity of Computation

Even though evidence sets are more complicated than standard fuzzy sets or IVFS, computationally they are still easier structures than general type 2 fuzzy sets or probabilistic sets. As discussed in section 2.3 IVFS require only an upper a lower value for their interval of membership, which is simpler than defining a probability or possibility distribution on membership degrees. Evidence sets are also interval-based set structures, thus each membership degree needs only to be described by simple intervals and their respective weight. If the number of intervals is kept fairly small, which is expected of human discriminative capacities, the complexity of computation is kept very small. In chapter 5 a computer application using evidence sets is proposed which faced no computational problems associated with the complexity of evidence sets. Indeed, the objects used to implement evidence sets are rather trivial compared to the larger relational database scheme in which they are imbedded.

4.4 Contextual Interpretation of Evidence Sets

> "To speak of *a prototype* at all is simply a convenient grammatical fiction; what is really referred to are *judgements* of degree of prototypicality." [Rosch, 1978, page 40 second italics added]

In the previous sections, the idea of categories as subjective creations of a cognitive agent doing the categorizing was stressed. It was also discussed how a full representation of uncertainty forms, as well as an adequate accounting of context are necessary attributes of good feasible models of categories. None of the fuzzy set and IVFS approaches to this problem consider, explicitly, the notion of subjective context dependencies. This is so because fuzzy sets do not offer an explicit account of belief in evidence; in other words, we have degrees of prototypicality and not *judgements* of degrees of prototypicality as Eleanor Rosch required in the previous quote.

The interpretation I suggest for the multiple intervals of evidence sets, in light of the problem of human categorization processes, considers each interval of membership $\boldsymbol{I}_j^x$, with its correspondent evidential weight $m^x(\boldsymbol{I}_j^x)$, as the representation of the prototypicality of a particular element *x* of *X* , in category *A* <u>*according*</u> to a particular *perspective*. In other words, each interval $\boldsymbol{I}_j^x$ represents a particular perspective of the element *x* of a category represented by an evidence set *A*. Thus, each element *x* of our evidence set *A* will have its membership varying within several intervals representing different, possibly conflicting, perspectives. An IVFS refers to the case where we have a single perspective on the category in question, even if it admits a nonspecific representation (an interval)[24].

The ability to maintain several of these perspectives, which may conflict at times, in representations of categories such as evidence sets, allows a model of cognitive categorization or knowledge representation to directly access particular contexts affecting the definition of a particular category, essential for radial categories. In other words, the several intervals of membership of evidence sets refer to different perspectives which explicitly point to particular contexts. In so doing, evidence sets facilitate the inclusion of subjectivity in models of cognitive categorization in addition to the inclusion of the several forms of uncertainty.

> "Whenever I write in this essay 'degree of support' that given evidence provides for a proposition or the 'degree of belief' that an individual accords the proposition, I picture in my mind an act of *judgement*. I do not pretend that there exists an objective relation between given evidence and a given proposition that determines a precise numerical degree of support. Nor do I pretend that an actual human being's state of mind with respect to a proposition can ever be described by a precise real number called his degree of belief, nor even that it can ever determine such a number. Rather, I merely suppose that an individual can make a *judgement*. Having surveyed the sometimes vague and sometimes confused perception and understanding that constitutes a given body of evidence, he can announce a number that represents the degree to which he judges that evidence to support a given proposition and, hence, the degree of belief he wishes to accord the proposition." [Shafer, 1976, p. 21, italics added]

Shafer's intent captured in the previous quotation seems to follow Rosch's earlier quotation in the context of cognitive categorization. The degrees of belief on which evidence theory is based do not aspire to be objective claims about some real evidence, they are rather proposed as *judgements*, formalized in the form of a degree. Likewise, Rosch's prototypes are not assumed to be an objective grading of concepts in a category, but rather judgements of some uncertain, highly context-dependent, grading. Evidence sets offer a way to model these ideas since an independent[25], unconstrained, membership grading of elements (concepts) in a category is offered together with an explicit formalization of the belief posited on this membership. In a sense, in evidence sets, membership in a category and judgments over membership are different, complementary, qualities of the same phenomenon. None of the other structures so far presented are able to offer both this independent characterization of membership and a formalization of judgments imposed on this membership: traditional set structures (crisp, fuzzy, or interval-valued) alone offer only an independent degree of membership, while evidence theory by itself offers primordially a formalization of belief which constrains the elements of a universal set with a probability restriction.

Regarding the previously discussed connectionist extraction of prototypes, notice that Evidence Sets, as any set structure, have independent, unconstrained membership. Connectionist prototypes are implicitly defined by a semantic metric constraining the elements of the categorizing universe. The existence of such metrics may be very important for cognitive categorization. However, and as previously stressed, Evidence Sets are merely proposed as models of cognitive categories, it is up to the model of cognitive categorization to supply additional constraints such as semantic metrics. As a higher level structure, it is very important that Evidence Sets do not have such constraints a priori, in fact, it is precisely their advantage over connectionist devices which are not flexible enough to allow users to arbitrarily change constraints and

---

[24]This idea of interpreting bodies of evidence as perspectives, spins off from a generalization of Gordon Pask's [1975, 1976] Conversation Theory which I have proposed with the construction of a data-retrieval system to be discussed in chapter 5 [Medina-Martins and Rocha, 1992].

[25]That is, every element of the set is formally free to be ascribed any value in the unit interval, independently of the values of other elements in the set.

contexts on prototype-based categories. Later in this chapter, approximate reasoning methods are proposed which shall be used in Chapter 5 to define a database retrieval system that constrains Evidence Sets with context-specific semantic metrics.

5. Relative Uncertainty and Evidence Sets

The nonspecificity and conflict of the membership degrees of evidence sets are defined on the unit interval which is uncountable. Thus, to measure the uncertainty content of evidence sets it is necessary to define measures of uncertainty capable of dealing with uncountable domains. Traditionally, uncertainty is measured in bits of information qualifying countable sets of alternatives. In the following I take a new approach to measure uncertainty-based information in uncountable domains.

5.1 Nonspecificity

5.1.1 A General Measure

Nonspecificity is identified with unspecified distinctions between several alternatives, that is, when several alternatives exist which are equally possible. It depends on the quantity of alternatives for the value or outcome of a particular proposition, event, variable, etc. Thus, to measure this kind of uncertainty, we need some measure of the "*size*" of the set of the several alternative values for a variable x. The "size" of several alternatives can be defined by a nonnegative, extended real valued set function on a σ-algebra $\mathscr{E}$ of X: μ: $\mathscr{E} \rightarrow [0, \infty]$. For intuitive reasons, it is desired that the "size" of a set of alternatives A, which contains a set of alternatives B plus some other elements, be larger than the "size" of B. Thus, μ must be *monotone*: $A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$, for all $A, B \in \mathscr{E}$. We further want that $\mu(\mathrm{A}) < \infty$ for some $A \in \mathscr{E}$, that is, there will be at least one element of $\mathscr{E}$ whose "size" is not infinite so that μ possesses at least some minimal discriminative power over $\mathscr{E}$. The set of all functions with the properties of μ is denoted by $\mathfrak{M}$.

A monotone set function μ will yield an intuitive measurement of the "size" of the several alternatives in an uncertain situation. However, in the context of DST, uncertain situations are described by bodies of evidence. Therefore, a measure of nonspecificity in DST should be a function NSp: $\mathscr{B} \times \mathfrak{M} \rightarrow [0, \infty)$ which takes into account the "size" of each focal element A, weighted by the respective probability assignment value $m(A)$:

$$NSp(m, \mu) = \sum_{A \subseteq X} m(A) \cdot \mu(A) \qquad \textbf{(11)}$$

Notice that since μ refers to set functions defined on a σ-algebra of X and not on $\mathscr{P}(X)$, the measures of uncertainty discussed below are only defined for bodies of evidence whose focal elements are members of a σ-algebra where μ can be defined. For instance, if μ is the Lebesgue measure, focal elements will be restricted to a maximal σ-algebra of Lebesgue-measurable sets.

Definition 1. Let f be a function defined on bodies of evidence of X, $f: \mathfrak{B} \rightarrow [0, \infty)$. f is *monotone* iff $f(m_1) \leq f(m_2)$, whenever $(\mathfrak{F}_1, m_1) \subseteq (\mathfrak{F}_2, m_2)$. $(\mathfrak{F}_1, m_1)$ and $(\mathfrak{F}_2, m_2)$ represent bodies of evidence defined on X [26].

Proposition 1. Function *NSp* is monotone as a function on $\mathfrak{B}$.

Proof: Dubois and Prade [1987] have shown that for any positive set-function λ such that if $A \subseteq B \Rightarrow \lambda(A) \leq \lambda(B)$, then

$$(\mathfrak{F}_1, m_1) \subseteq (\mathfrak{F}_2, m_2) \Rightarrow \sum_{A \subseteq X} m_1(A) \cdot \lambda(A) \leq \sum_{A \subseteq X} m_2(A) \cdot \lambda(A)$$

Since μ in (11) is a monotone positive set function, the above implication holds. If we substitute λ for μ it follows that:

$$(\mathfrak{F}_1, m_1) \subseteq (\mathfrak{F}_2, m_2) \Rightarrow NSp(m_1, \mu) \leq NSp(m_2, \mu)$$

□

5.1.2 Absolute Nonspecificity

Monotonicity alone is clearly not enough for what we desire of measures of nonspecificity. Traditionally, measures of uncertainty observe other properties such as additivity and subadditivity.

Definition 2. Let f be a function defined on the set of all bodies of evidence, $f: \mathfrak{B} \rightarrow [0, \infty)$. Let m be an arbitrary joint basic probability assignment defined on $\mathcal{P}(X \times Y)$, and m_X, m_Y be the associated marginal basic probability assignments. f is *subadditive* iff $f(m) \leq f(m_X) + f(m_Y)$.

Definition 3. Given the same conditions as in definition 2, except that m is now based on <u>noninteractive bodies of evidence</u>. f is *additive* iff $f(m) = f(m_X) + f(m_Y)$.

Proposition 2. $NSp(m, \mu_{X\times Y})$ as defined in (11), is additive if and only if there exist monotone functions μ_X, and $\mu_Y \in \mathfrak{M}$, defined on the sets X, and Y, which satisfy:

$$\mu_{X\times Y}(R) = \mu_X(R_X) + \mu_Y(R_Y) \qquad \textbf{(12a)}$$

when $R = R_X \times R_Y$ (noninteractive bodies of evidence), where R_X and R_Y are the projections of R on X and Y respectively as defined in section 2.2.4, and we have either:

$$\mu_{X\times Y}(R) \leq \mu_X(R_X) + \mu_Y(R_Y) \qquad \textbf{(12b)}$$

or

$$\mu_{X\times Y}(R) \geq \mu_X(R_X) + \mu_Y(R_Y) \qquad \textbf{(12c)}$$

for all cases $(R \subseteq R_X \times R_Y)$.

Proof: Additivity is defined for noninteractive bodies of evidence. Utilizing (8), (12a), and (3) we reach:

---

[26] Inclusion in DST was discussed in section 2.2.6.

$$
\begin{aligned}
NSp(m,\mu_{X\times Y}) &= \sum_{R\subseteq X\times Y} m(R)\cdot\mu_{X\times Y}(R) = \\
&= \sum_{A\subseteq X|A=R_X}\ \sum_{B\subseteq Y|B=R_Y} m_X(A)\cdot m_Y(B)\cdot\left[\mu_X(A)+\mu_Y(B)\right] = \\
&= \sum_{A\subseteq X}\ \sum_{B\subseteq Y} m_X(A)\cdot m_Y(B)\cdot\mu_X(A) + \sum_{A\subseteq X}\ \sum_{B\subseteq Y} m_X(A)\cdot m_Y(B)\cdot\mu_Y(B) = \\
&= \sum_{A\subseteq X} m_X(A)\cdot\mu_X(A)\cdot\sum_{B\subseteq Y} m_Y(B) + \sum_{B\subseteq Y} m_Y(B)\cdot\mu_Y(B)\cdot\sum_{A\subseteq X} m_X(A) = \\
&= NSp(m_X,\mu_X) + NSp(m_Y,\mu_Y)
\end{aligned}
$$

The reverse implication is true if (12b) or (12c) is also satisfied. If we multiply $NSp(m_X, \mu_X)$ and $NSp(m_Y, \mu_Y)$ by the unitary quantities given by (3) defined on sets Y and X respectively, and arrange the terms in the same manner above, we conclude that $\mu_X(A) + \mu_Y(B)$ must equal $\mu_{X\times Y}(A \times B)$,if we are to recover $NSp(m, \mu_{X\times Y})$, only if (12b) or (12c), since $\sum_i a_i b_i = \sum_i a_i c_i \Rightarrow b_i = c_i,\ iff\ \forall_i$ either $b_i - c_i \geq 0$ or $b_i - c_i \leq 0$

□

Lemma 1. Let R_X (R_Y) represent the projection of the elements R of $Z=X \times Y$ on X (Y). m_X (m_Y) is the marginal basic probability assignments defined on X (Y). m is the joint basic probability assignment on Z. Then the following equations holds:

$$NSp(m_X,\mu_X) = \sum_{R\subseteq X\times Y} m(R)\cdot\mu_X(R_X)$$

Proof:

$$
\begin{aligned}
NSp(m_X,\mu_X) &= \sum_{A\subseteq X} m_X(A)\cdot\mu_X(A) = \sum_{A\subseteq X}\left(\sum_{R|A=R_X} m(R)\right)\cdot\mu_X(A) \\
&= \sum_{A\subseteq X}\ \sum_{R|A=R_X} m(R)\cdot\mu_X(R_X) = \sum_{R\subseteq X\times Y} m(R)\cdot\mu_X(R_X)
\end{aligned}
$$

□

Proposition 3. If $NSp(m, \mu_{X\times Y})$ is additive for noninteractive bodies of evidence, then if (12b) is satisfied it is subadditive for interactive ones.

Proof: If NSp is additive then there exist monotone functions μ_X, and $\mu_Y \in \mathfrak{M}$ defined on X and Y, respectively, which satisfy (12). Therefore, using lemma 1 and (12b) for the projections of R on X and Y:

$$
\begin{aligned}
NSp(m_X,\mu_X) + NSp(m_Y,\mu_Y) &= \sum_{R\subseteq X\times Y} m(R)\cdot\left[\mu_X(R_X) + \mu_Y(R_Y)\right] \\
&\geq \sum_{R\subseteq X\times Y} m(R)\cdot\mu_{X\times Y}(R) = NSp(m,\mu_{X\times Y})
\end{aligned}
$$

□

When X is countable, a possible set function μ_X is the Hartley[1928] measure:

$$\mu_X(A) = \log_2|A|, \ \forall A \subseteq X \qquad \textbf{(13)}$$

where $|A|$ denotes the cardinality of set A, which yields an intuitive value of zero for no uncertainty (unitary cardinality). Substituting (13) into *NSp* (11), we obtain the familiar measure of nonspecificity, *N*, that Dubois and Prade [1985] generalized from Higashi and Klir's [1983] *U-uncertainty*:

$$NSp(m,\mu_X) = N(m) = \sum_{A \subseteq X} m(A) \cdot \log_2|A| \qquad \textbf{(14)}$$

which attains the value 0 for full certainty, and $\log_2|X|$ for total ignorance. As a special case of *NSp* (11), and since μ_X satisfies (12a) and (12b), *N* is monotone, additive, and subadditive. In addition, Ramer[1987] showed that *N* expressed in (14) is unique under a set of desirable axioms which include the three axioms above, plus others such as *symmetry*, *branching*, and *normalization* (based on choosing bits as measurement units) [see also Klir, 1993 for more details].

All of these characteristics made this measure of nonspecificity the obvious choice for measuring nonspecificity in countable domains. However, its extension to uncountable, domains introduces a few problems that I shall investigate next. Before that though, notice that μ_X given by (13) is not defined for the empty set, since $|\varnothing|=0$. This is a technicality that can be avoided by rewriting (14) as a sum for all $A \subseteq \mathfrak{F}$ [Klir, 1993], where $\mathfrak{F}$ is the set of focal elements, which does not include the empty set by definition.

If the domain of μ , *Y*, is uncountable[27], μ_Y given by (13) clearly yields undesirable values. Only discrete, finite subsets *B* of *Y* will have $\mu_Y(B) < \infty$, since the cardinality of all other subsets of *Y* will be infinite. Therefore, all bodies of evidence defined on infinite, uncountable, domains whose focal elements are not finite, will have infinite nonspecificity as defined by (14). To improve this situation, a natural extension of (13) can be found by replacing cardinalities by lengths, or better, by a *Lebesgue measure* $\lambda(B)$ of the subsets *B* of *Y*. In general we can use the symbol $\overline{B}$ (length of *B*) to denote the Lebesgue measure of *B*. The new μ_Y comes:

$$\mu_Y(B) = \ln(\overline{B}), \text{ whenever } \overline{B} \text{ makes sense} \qquad \textbf{(15)}$$

The natural logarithm is used instead of the logarithm of base 2, since the choice of the logarithm of base 2 in (13) is motivated solely by the desire to employ bits as measurement units. In uncountable domains, the notion of bit is meaningless since we have an infinite universe of alternatives. Substituting (15) into *NSp* (11) we obtain:

$$NSp(m,\mu_Y) = N(m) = \sum_{B \in \mathscr{F}} m(B) \cdot \ln(\overline{B}) \qquad \textbf{(16)}$$

This measure of nonspecificity can also be seen to follow from Ramer's [1990] work on measuring information in infinite domains. Since μ_Y given by (15) satisfies (12) (with important limitations to be discussed below), we can say that *N* as defined in (16) is monotone, additive, and subadditive <u>whenever μ_Y is defined and satisfies (12)</u>. Now, μ_Y in (15) must be a positive set function as previously required[28], thus it is only defined for subsets *B* of *Y* with Lebesgue measure $\lambda(B) \geq 1$. This can be avoided by defining it

[27] Note that the bodies of evidence considered here, and throughout the paper, have a finite number of focal elements (see section 2.2.3).

[28] Since we desire positive measures of information.

instead as: $\mu_Y(B) = \ln(1+\overline{B})$. However, clearly, this alternative definition does not satisfy (12a); therefore, it would lead to a definition of nonspecificity without the desired axiomatic requirements.

Notice that if Y is the unit interval, an important domain for interval computation, we can redefine μ as: $\mu^*_Y(B) = -\ln(\overline{B})$, a positive set function, though inversely monotonic[29] , which satisfies (12a) and (12c). Substituting this expression into the general *NSp* (11) we obtain the inverse of N in (16): $N^* = -N$, which is additive and subadditive for any focal element with positive Lebesgue measure. Because of its inverse monotonicity, it intuitively measures the inverse of nonspecificity: *specificity*. Turksen [1994] has proposed the inverse of specificity as a measure of nonspecificity for interval valued fuzzy sets, however, since $\mu_Y(B) = -1/\ln(\overline{B})$ does not satisfy (12a), such a measure of nonspecificity is not additive.

The measure of specificity for the unit interval N^*, is clearly more effective than N in (16) itself, since the latter is restricted to bodies of evidence with focal elements with Lebesgue measure greater or equal than one. Further, bounded uncountable, set can always be mapped into the unit interval, which makes N^* applicable to any infinite, uncountable domain Y. However, μ^*_Y is not defined for countable subsets of Y, since the Lebesgue measure of subsets B of Y with countably many elements is null. Thus, specificity is only defined for subsets of Y with uncountably many elements. This leaves out any body of evidence containing a countable collection of singletons as a focal element, that is, any zero Lebesgue set. For instance, full certainty cannot be measured since it is defined by a body of evidence with a singleton as the only focal element.

Another point worth mentioning regarding this formulation of measures of nonspecificity has to do with the utilization of the set function μ in different domains. Notice that in (12), μ must be defined differently for $X \times Y$, X, and Y: the relevant universal sets necessary to define additivity and subadditivity. A function is a procedure to calculate a relation between two sets: a domain and a range. If the domain changes, so does the definition of the function. When we use the same function on different domains, it means that there exists a larger set U, containing all the different domains, where the function is actually defined. For instance, we use the same expression (14) to measure the nonspecificity N of bodies of evidence defined on countable sets $X \times Y$, X, and Y based on the same function μ given by (13). In other words, we can calculate cardinalities in all those sets because they are countable. Cardinality is a measure that is defined for the set of all countable sets, therefore we can use the same procedure as we move from countable domain to countable domain.

In uncountable domains the situation changes since the Lebesgue measure on $X \times Y$ is different from the Lebesgue measure on X or Y. Basically, it is an area for the first case, and a length for the second. The Lebesgue measure is defined in $\mathbb{R}^n$ as a general procedure. This allows us to move from one uncountable domain to another and still be able to calculate the Lebesgue measure. However, we should be aware that unlike cardinality, the Lebesgue measure is calculated in a different way as we move from sets to the cross products of sets.

This caution becomes more relevant if we wish to measure the uncertainty of a body of evidence defined on $X \times Y$, when X is countable, and Y is uncountable: *hybrid domains*. The nonspecificity measures N given by (14) or (16) (as well as the related specificity N^*) cannot deal with such a case because they are committed to the cardinality or the Lebesgue measure of a set, respectively, for any domain they are applied to. However, the general measure of nonspecificity *NSp* (11) works for such a hybrid case. Furthermore, provided we define $\mu_{X\times Y}$ using (12), where μ_X and μ_Y are given, for instance, by (13) and (15) respectively, *NSp* will be additive and subadditive (propositions 2 and 3). Of course (15) carries the limitations discussed above for uncountable domains yet to be improved if we are to treat the information content of uncountable and hybrid domains effectively.

---

[29] $\mu(A) \geq \mu(B)$ when $A \subseteq B$.

The reason why I refer to N in (13) and (16) as a measure of *absolute nonspecificity*, will become clearer in the next section with the introduction of *relative nonspecificity*. Basically, N measures the nonspecificity of bodies of evidence regardless of the amount of maximum uncertainty-based information present in their universal sets. A body of evidence will carry the same amount of nonspecificity whether it is defined on X, or on some set $U \supseteq X$. In countable domains N's (14) unit is the bit (1 bit equals the uncertainty of 2 alternatives), while in uncountable domains N's (16) unit is associated with the natural number e: when the Lebesgue measure of the set of alternatives equals e, nonspecificity is unitary.

Let me now summarize the problems facing N defined by (14) and (16):

a. (14) is only defined for bodies of evidence with countable, non-empty, focal elements defined on some universal set X.
b. (16) is only defined for bodies of evidence with focal elements whose Lebesgue measure is greater than or equal to one, while the related specificity is defined only for focal elements with positive Lebesgue measure. This excludes singletons as focal elements.
c. Neither (14) nor (16) alone can deal with hybrid domains defined as cross-products of countable and uncountable sets.
d. The general measure (11), with μ satisfying by (12) can deal with hybrid domains and it is monotone, additive, and subadditive. Nonetheless, if μ is based upon (13) and (15), *NSp* will be restricted in the same manner as (14) or (16).

Notice that the above discussion does not include a very recent development in the measurement of nonspecificity in nondiscrte domains. Klir and Yuan [1995] have proposed a Hartley-like function more complex than (15) with much better axiomatic properties than (16) and which solves the problems in b and c above. A more complete study of this new function in the framework here proposed is left for future research. In any case, one of the reasons to develop the measures of relative nonspecificity to be presented in the next section is the establishment of more computationally friendly measurements of nonspecificity, which are simpler than the new Hartley-like function.

5.1.2 Relative Nonspecificity

The measures of nonspecificity obtained so far from the general *NSp* defined by (11), are based on a generic monotone set function μ. I will now restrict μ by making it a classic measure with properties (1) and (2) defined in section 2.1. Further, instead of (12) used in 5.1.1 for the measures of absolute nonspecificity, let it satisfy the following restriction for domains defined by the Cartesian product of two sets X and Y:

$$\mu_{X\times Y}(R) = \mu_X(R_X)\cdot\mu_Y(R_Y) \qquad \textbf{(17a)}$$

when $R = R_X \times R_Y$ (noninteractive bodies of evidence), where R_X and R_Y are the projections of R on X and Y respectively as defined in section 2.2.5, and we have:

$$\mu_{X\times Y}(R) \leq \mu_X\left(R_X\right)\cdot\mu_Y\left(R_Y\right) \qquad \textbf{(17b)}$$

for all cases ($R \subseteq R_X \times R_Y$). If X is countable $\mu_X(A)$ can be the *cardinality* of set A: $\mu_X(A) = |A|$. If Y is uncountable, $\mu_Y(B)$ can be the *length* (Lebesgue measure) of set B: $\mu_Y(B) = \overline{B}$.

Proposition 4. If μ in $NSp(m, \mu)$ given by (11) follows (17), then NSp satisfies the following *multiplicative property* for noninteractive bodies of evidence $(\mathcal{F}_X, m_X)$ and $(\mathcal{F}_Y, m_Y)$:

$$NSp(m, \mu_{X\times Y}) = NSp(m_X, \mu_X) \cdot NSp(m_Y, \mu_Y) \qquad \textbf{(18)}$$

Proof.

$$\begin{aligned} NSp(m, \mu_{X\times Y}) &= \sum_{A\times B\subseteq X\times Y} m(A\times B) \cdot \mu_{X\times Y}(A\times B) = \\ &= \sum_{A\times B\subseteq X\times Y} m_X(A) \cdot m_Y(B) \cdot \mu_X(A) \cdot \mu_Y(B) = \\ &= \sum_{A\subseteq X} m_X(A) \cdot \mu_X(A) \cdot \sum_{B\subseteq Y} m_Y(B) \cdot \mu_Y(B) = \\ &= NSp(m_X, \mu_X) \cdot NSp(m_Y, \mu_Y) \end{aligned}$$

Utilizing (8) and (17a).
□

NSp with μ following (17) is thus monotone and observes multiplicative behavior for noninteractive bodies of evidence. In countable domains, with $\mu(A) = |A|$, it varies between 1 for full certainty, and $|X|$ for total ignorance. In uncountable domains, with $\mu(A) = \overline{A}$, *NSp* varies between 0 for full certainty, and $\overline{X}$ for total ignorance. However, even though it has a few intuitive properties which allow us to measure nonspecificity fairly well, it lacks some other desirable axiomatic properties. To improve these, I will impose another restriction on μ with implications for the modeling of nonspecificity discussed ahead.

Definition 4. Let μ be a classical measure on $\mathcal{P}(X)$ taking values on the unit interval: μ: $\mathcal{P}(X) \to [0, 1]$.

In other words, μ, as a measure, follows (1) and (2), is further restricted by (17), and <u>it yields values $\leq$ 1</u>.

Proposition 5. $NSp(m, \mu)$ given by (11), with μ following definition 4, and restricted by (17), is subadditive.

Proof: Functions $\mu_{X\times Y}$, μ_X, and μ_Y defined on $X \times Y$, X, and Y respectively satisfy (17). Using lemma 1,

$$\begin{aligned} NSp(m_X, \mu_X) + NSp(m_Y, \mu_Y) &= \sum_{R\subseteq X\times Y} m(R) \cdot \left[\mu_X(R_X) + \mu_Y(R_Y)\right] \\ &\geq \sum_{R\subseteq X\times Y} m(R) \cdot \mu_X(R_X) \cdot \mu_Y(R_Y) \end{aligned}$$

Since the sum of two numbers in the unit interval is always larger or equal than their product, then: $\mu_X(R_X) + \mu_Y(R_Y) \geq \mu_X(_X) \cdot \mu_Y(R_Y)$. Therefore, using (17b):

$$\sum_{R\subseteq X\times Y} m(R) \cdot \mu_X(R_X) \cdot \mu_Y(R_Y) \geq \sum_{R\subseteq X\times Y} m(R) \cdot \mu_{X\times Y}(R) = NSp(m, \mu)$$

thus $NSp(m_X) + NSp(m_Y) \geq NSp(m)$.
□

Thus, *NSp* given by (11), with μ following definition 4, and restricted by (17) is monotone, subadditive, and observes a multiplicative property for noninteractive bodies of evidence.

When X is countable, an obvious choice for the measure μ_X is the cardinality of a set divided by the cardinality of the universal set:

$$\mu_X(A) = \frac{|A|}{|X|},\ \forall A \subseteq X \tag{19}$$

Substituting (19) into *NSp* (11) we obtain a measure of nonspecificity *IN* given by:

$$NSp(m_X, \mu_X) = IN(m_X) = \frac{1}{|X|} \sum_{A \subseteq X} m_X(A) \cdot |A| \tag{20}$$

The maximum value it attains for total ignorance is 1, and the minimum for full certainty is 0. It is defined for all subsets of X, including the empty set. It is clearly symmetrical in the sense that bodies of evidence with the same distribution of evidential weights, applied to focal elements of equal cardinality, will yield the same value of nonspecificity. It is monotone, subadditive, and multiplicative for noninteractive bodies of evidence, for all $A \in \mathscr{P}(X)$ (propositions 1, 5, and 4 respectively). It is a ratio, thus it is unitless[30].

When Y is uncountable, the measure μ can be the Lebesgue measure, or length in the case of a one dimensional domain, of a set divided by the Lebesgue measure of the universal set:

$$\mu_Y(B) = \frac{\overline{B}}{\overline{Y}},\ \forall B \subseteq Y \tag{21}$$

which can only be applied to universal sets Y with finite Lebesgue measure greater than zero. Substituting (21) into *NSp* (11) we obtain a measure of nonspecificity for uncountable, domains:

$$NSp(m_Y, \mu_Y) = IN(m_Y) = \frac{1}{\overline{Y}} \sum_{B \subseteq Y} m_Y(B) \cdot \overline{B} \tag{22}$$

The maximum value it attains for total ignorance is 1, and the minimum for full certainty is 0. It is defined for all subsets B of Y, including zero Lebesgue sets, and the empty set. It is also symmetrical, though in this case symmetry has to be defined in terms of the Lebesgue measure not cardinality. It is monotone, subadditive, and multiplicative for noninteractive bodies of evidence, for all $B \in \mathscr{P}(Y)$. It is unitless.

IN given by (20) and (22) cannot deal with hybrid domains because they are committed to specific measures μ_X (19) and μ_Y (21) respectively. However, the general expression *NSp* (11) can measure domains defined as the cross-product of countable and uncountable sets, provided we use restriction (17) to define

[30] Nonspecificity measure (20) is in some ways similar to Yager's [1982, 1983] specificity. However, Yager's measure deals with the intuitive inverse of nonspecificity, but in absolute terms. Even though it is restricted to the unit interval, the specificity content of a body of evidence is independent on the size of its universal set.

the joint measure $\mu_{X\times Y}$, and, for instance, define μ_X with (19), and μ_Y with (21), all following definition 4. Such a measure is monotone, multiplicative, subadditive, and symmetrical.

The measure of nonspecificity $IN(m)$ for countable and uncountable domains, given by (20) and (22) respectively, can be seen as an *index of nonspecificity*. It accounts nonspecificity by relating it to the maximum information present in the universal set (measured by the respective μ): *relative nonspecificity*. This is in many ways a more intuitive way to measure nonspecificity than utilizing absolute measures such as N (14) and (16). When we measure the nonspecificity of a body of evidence in absolute terms, say in bits as yielded by (14), we know how many options we have: 1 bit, 2 options. Nonetheless, 2 options out of 10 possible options do not quite mean the same thing as 2 options out of 100. The first case clearly represents a more (relative) uncertain situation. Measures such as IN offer this kind of relative to the universal set measurement of nonspecificity, which might be more intuitive in practical situations such as reliable computation [Rocha et al, 1996]. Further, IN (22) is clearly much more effective than N (16) in measuring nonspecificity in infinite, uncountable, domains. It is defined for any kind of focal element on infinite, uncountable, domains, including zero Lebesgue sets; (16) cannot treat those. Also, even in countable domains, (20) is defined on $\mathcal{P}(X)$, the provision to exclude the empty set is not necessary. Within this framework, the general measure NSp (11) can be used to deal with hybrid domains effectively as discussed above. Relative nonspecificity, with its unitless measures, offers intuitively a more coherent framework to measure uncertainty in countable and uncountable domains.

5.2 Conflict

Conflict is identified with disagreement between several alternatives. The word disagreement implies the existence of some distinctive criteria between the several alternatives. When it is possible to distinguish between the several alternatives for some event, proposition, or variable values, we have conflict amongst alternatives. It depends on the strength of the several alternatives, not on the "size" of the sets of distinguishable alternatives. Conflict increases when the quantity of equally strong contending alternatives increases, and decreases otherwise.

5.2.1 Absolute Conflict

In probability theory, conflict is measured with the Shannon measure of entropy: $H(m) = -\sum m(\{x\})\log_2 m(\{x\})$. In evidence theory, the probabilistic entropy is efficiently extended, on countable domains, to the measure of uncertainty $S(m)$ named *Strife* given by equation (9) in section 2.4.1. It measures the mean conflict among evidential claims within each given body of evidence in bits. It is additive, and it becomes the Shannon entropy for probability measures [Klir, 1993]. Its range is $[0, \log_2 n]$, where n is the number of focal elements: $n = |\mathfrak{F}|$. $S(m) = 0$ when $m(\{x\}) = 1$ for some x of X, and $S(m) = \log_2 n$ when $m(A) = 1/n$ for all $A \in \mathfrak{F}$, provided all focal elements A are disjoint. Thus, strife increases as alternatives start losing their distinctive strength and are qualified by a similar belief weight m.

The term $|A \cap B|/|A|$ in (9) above, even though utilizing the cardinality measure, is simply expressing the *degree of subsethood* of set A in set B [Klir, 1993]. Conflict in evidence theory does not depend on the “size” of the several focal elements $A \in \mathfrak{F}$, but on their strength, given by the basic probability assignment m, and on the degree of subsethood in one another. Therefore, since $\mathfrak{F}$ is finite by definition (section 2.2.3), as long as a suitable degree of subsethood is defined, the measure of strife given by (9) remains an adequate measure of conflict (in bits and with the same characteristics above), both for countable and for uncountable domains:

$$S(m) = -\sum_{A \in \mathcal{F}} m(A) \log_2 \sum_{B \in \mathcal{F}} m(B) \cdot \mathrm{SUB}(A, B) \tag{23}$$

where SUB(A, B) denotes the subsethood of set A in set B of X. In countable domains its value is given by the ratio $|A \cap B| / |A|$. In uncountable domains it is given by:

$$\mathrm{SUB}(A,B) = \begin{cases} 1, & \text{if } A \subseteq B \\ \dfrac{\lambda(A \cap B)}{\lambda(A)}, & \text{if } \lambda(A) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{24}$$

where $\lambda(A)$ represents the Lebesgue measure of set A. The measure of discord as discussed in section 2.4, can in the same way be easily extended to nondiscrete domains.

5.2.2 Relative Conflict

Strife given by (23) measures conflict in uncertain situations, described in DST, in absolute terms. That is, it is a measure of the mean conflict among evidential claims within each given body of evidence. As with the case of nonspecificity, we may be interested in a relative measure of conflict, relating the amount of conflict present on a specific situation to the maximum conflict present in a given body of evidence with n focal elements:

$$IS(m) = \frac{S(m)}{\log_2 n} \tag{25}$$

IS is an *index of conflict* yielding a value in the unit interval. It attains the value 1 for maximum conflict, and 0 for no uncertainty. *IS* is not additive quite in the same sens
e as *S*, but it obeys the following proposition:

Proposition 6. $IS(m)$ given by (25) satisfies the following property for noninteractive bodies of evidence $(\mathfrak{F}_X, m_X)$ and $(\mathfrak{F}_Y, m_Y)$:

$$IS(m) = w_1 \cdot IS(m_X) + w_2 \cdot IS(m_Y), \text{ with } w_1 + w_2 = 1 \tag{26}$$

where, m is the basic probability assignment of a body of evidence defined on $Z = X \times Y$, and $w_1 = \log n_X / (\log n_X + \log n_Y)$, and $w_2 = \log n_Y / (\log n_X + \log n_Y)$; n_X and n_Y denote the number of focal elements in $\mathfrak{F}_X$ and $\mathfrak{F}_Y$ respectively.

The proof is quite trivial and follows from the fact that *S* is additive for noninteractive bodies of evidence.

Proposition 6 indicates that the index of conflict for a joint body of evidence defined on the Cartesian product of two sets, is a weighted sum of the indices of conflict of each marginal body of evidence. The weights are a measure of the relative size of the marginal bodies of evidence, that is, the number of focal elements of each marginal body of evidence. The weighting is necessary to maintain the notion of index of conflict. For interactive bodies of evidence, the equality in (26) is substituted by an inequality "$\leq$", which defines a kind of weighted subadditivity.

5.3 Fuzziness

The amount of fuzziness present in a fuzzy set was defined in section 2.4.3 as the lack of distinction between the set and its complement, and given by equation (11). It is fairly easy to extend the usual fuzzy set operations of complement, intersection, and union to an IVFS framework [e.g. Gorzałczany, 1987] or, more generally, to an evidence set framework. Section 6 is precisely devoted to the definition of such an extended theory of approximate reasoning [an equivalent formulation has also been proposed by Zhu and Lee, 1995].

The interval valued membership function of elements of X in an IVFS A is given by: $A(x) = I^x = [l^x_{inf}, l^x_{sup}] \subseteq [0,1]$. Its complement can be defined as the negation of the interval limits in reverse order: $A^c(x) = (I^x)^c = [1 - l^x_{sup}, 1 - l^x_{inf}]$. The membership function of an evidence set A of X is given, for each x, by n intervals weighted by a basic probability assignment m^x:

$$A(x) = \left\{ \left\langle I_1^x, m_1^x \right\rangle, \left\langle I_2^x, m_2^x \right\rangle, \cdots, \left\langle I_n^x, m_n^x \right\rangle \right\} \tag{27}$$

The complement of an evidence set, or the negation operator in interval valued evidential logic systems [Zhu and Lee, 1995], is defined as the complement of each of its interval focal elements with the preservation of their respective evidential strengths:

$$A^c(x) = \left\{ \left\langle \left(I_1^x\right)^c, m_1^x \right\rangle, \left\langle \left(I_2^x\right)^c, m_2^x \right\rangle, \cdots, \left\langle \left(I_n^x\right)^c, m_n^x \right\rangle \right\} \tag{28}$$

Since the complementation of a set does not affect the Lebesgue measure of the interval focal elements, nor their evidential weight, the distinction, D, between an interval I^x and its complement $(I^x)^c$, is defined as the absolute of the difference between their respective higher bounds:

$$D\left(I^x, \left(I^x\right)^c\right) = \left| l^x_{sup} - \left(l^x_{sup}\right)^c \right| = \left| l^x_{sup} - (1 - l^x_{inf}) \right| = \left| l^x_{sup} + l^x_{inf} - 1 \right| \tag{29}$$

Notice that all intervals symmetric to the middle point (½) of the membership space $Y = [0, 1]$, will be indistinguishable from their complement ($D = 0$) since the sum of their limits is 1.

The fuzziness of the membership of an element x of X in an evidence set A, is obtained by weighting the quantity D for each focal element with m^x, and subtracting it from 1 to obtain the *lack* of distinction between its membership in the set and in its complement (such quantity is an index of local fuzziness):

$$IF^x(A(x)) = 1 - \sum_{k=1}^{n} m^x\left(I_k^x\right) \cdot D\left(I_k^x, \left(I_k^x\right)^c\right) \tag{30}$$

The absolute measure of fuzziness for the whole evidence set A is:

$$F(A) = \sum_{x \in X} IF^{x}\big(A(x)\big) \tag{31}$$

whose range is [0, $|X|$]. An index of total relative fuzziness can be obtained by:

$$IF(A) = \frac{F(A)}{|X|} \tag{32}$$

Notice that if X is uncountable, (31) and (32) can de adapted so that the sum becomes an integral, and the cardinality of X becomes the Lebesgue measure of X.

5.4 3-D Uncertainty

A fuzzy set captures fuzziness in a specific way; an IVFS introduces nonspecificity; a consonant evidence set introduces grades or shades of nonspecificity; and finally, a nonconsonant evidence set introduces conflict as we have cases where the degree to which an element is a member of a set is represented by disjoint sub-intervals of [0, 1] with different evidential strengths. The three forms of uncertainty are clearly present in human cognitive processes. As exposed above, fuzzy sets and interval valued fuzzy sets offer only a limited representation of recognized uncertainty forms, while evidence sets capture all of those. Thus, more than simply measuring fuzziness, as approximate reasoning models do, models of uncertain reasoning based on evidence sets need to effectively measure all the three uncertainty forms. Hence, we need a 3-tuple of measures of the 3 main kinds of uncertainty to aid us in the decision making steps of our uncertain reasoning models. Each situation, each set, should be qualified in its uncertainty content with something like: (*Fuzziness*, *Nonspecificity*, *Conflict*).

F and IF, equations (31) and (32), presented in section 5.3 define, respectively, measures of absolute and relative fuzziness for evidence sets. Given the results of sections 5.1 and 5.2, we can now define similar measures for nonspecificity and conflict.

5.4.1 Nonspecificity in Evidence Sets

In addition to fuzziness, the membership of an element x of X in an evidence set A possesses nonspecificity that can be measured by IN (22) derived in section 5.1.2 for uncountable domains such as Y=[0, 1]. This measure defines an index of local relative nonspecificity IN^{x} ($\overline{Y}$ = 1):

$$IN^{x}(A(x)) = \sum_{I^{x} \subseteq [0,1]} m^{x}(I^{x}) \cdot \overline{I^{x}} \tag{33}$$

An absolute measure of nonspecificity for the whole evidence set A is:

$$N(A) = \sum_{x \in X} IN^{x}\big(A(x)\big) \tag{34}$$

whose range is [0, $|X|$]. An index of total relative nonspecificity can be obtained by:

$$IN(A) = \frac{N(A)}{|X|} \tag{35}$$

Again, if X is uncountable, (34) and (35) can de adapted so that the sum becomes an integral, and the cardinality of X becomes the Lebesgue measure of X.

5.4.2 Conflict in Evidence Sets

In addition to fuzziness and nonspecificity, the membership of an element x of X in an evidence set A possesses conflict that can be measured by IS (25) derived in section 5.2. This measure defines an index of local relative conflict:

$$IS^{x}(A(x)) = \frac{S(m^{x})}{\log_2 n} \tag{36}$$

where $S(m^x)$ is the measure of strife (23) derived in 5.2, now defined on a body of evidence ($\mathfrak{F}^x$, m^x) on Y=[0, 1], where $\mathfrak{F}^x$ is the set of focal elements I^x, and n is the number of such focal elements. An absolute measure of conflict for the whole evidence set A is:

$$S(A) = \sum_{x \in X} IS^{x}(A(x)) \tag{37}$$

whose range is [0, $|X|$]. An index of total relative conflict can be obtained by:

$$IS(A) = \frac{S(A)}{|X|} \tag{38}$$

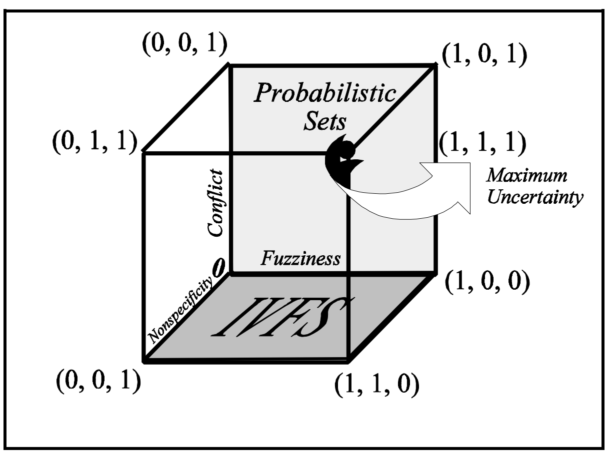

Figure 4: 3-D Uncertainty Cube

Once more, if X is uncountable, (37) and (38) can de adapted so that the sum becomes an integral, and the cardinality of X becomes the Lebesgue measure of X.

5.4.3 3-D Uncertainty Cube

The three forms of uncertainty define a 3 dimensional uncertainty space for set structures, where crisp sets occupy the origin, fuzzy sets the fuzziness axis, IVFS the fuzziness-nonspecificity plane, and evidence sets most of the rest of this space. Probabilistic sets occupy the conflict-fuzziness plane. Notice that evidence sets cannot occupy the conflict-nonspecificity plane, that is, if nonspecificity and conflict exist in evidence sets, then so will fuzziness. If the measures of uncertainty used are the indices so far presented, the uncertainty content of evidence sets can be described in a unit cube space (figure 4). We can calculate the uncertainty of each element x of X of an evidence set A (local uncertainties), and plot each element in this cube, in which case the uncertainty of an evidence set will be described by a "cloud" of points in the uncertainty cube. The local uncertainty, $\boldsymbol{U}^{x}$, will be defined by the following (fuzziness, nonspecificity, conflict) 3-tuple based on equations (30), (33), and (36) respectively:

$$U^{x}(A(x)) = \left(IF^{x}(A(x)), IN^{x}(A(x)), IS^{x}(A(x))\right) \quad \textbf{(39)}$$

Alternatively, we may calculate the total uncertainty indices for the set A, and plot it in this cube as a single point. Naturally, this point will be the center of mass of the cloud of local uncertainties. The total uncertainty, $\boldsymbol{U}$, of an evidence set A will be defined by:

$$U(A) = \left(IF(A), IN(A), IS(A)\right) \quad \textbf{(40)}$$

The uncertainty situation of the several set structures known is summarized in the following table:

Uncertainty situation of different set structures

	U		
	IF	*IN*	*IS*
Crisp Sets	0	0	0
Fuzzy Sets	[0,1]	0	0
IVFS	[0,1]	[0,1]	0
Evidence Sets	[0,1]	[0,1]	[0,1]

6. Belief-Constrained Approximate Reasoning

6.1. Uncertainty Increasing Operations Between Evidence Sets

Recently, Zhu and Lee [1995] have proposed a belief based multi valued logic which defines a connection between evidence theory and multi valued logics in much of the same way as evidence sets do, that is, with the establishment of degrees of belief on truth values given by intervals of the unit interval. While evidence sets were defined in the context of set theory, Zhu and Lee thought of this extension in terms of multi valued logics. This way, in the former we speak of belief based, interval valued membership in a set, while in the latter we speak of belief based, interval valued truth value of a proposition. Most of the operators discussed in this section are equivalent to Zhu and Lee's formulation, though their interpretation might differ.

The operations of complementation, intersection, and union are the most basic connectives in a theory of approximate reasoning. here I discuss only these operators, since all other connectives can be easily constructed from these. Naturally, complementation, intersection, and union as defined below for evidence sets, subsume as special cases, the same operations for IVFS and fuzzy sets.

6.1.1 Complementation

The complement of an evidence was already defined in section 5.3 with equation (28).

6.1.2 Intersection

The intersection of two IVFS [Gorzałczany, 1987] is defined as the minimum of their respective lower and upper bounds of their membership intervals. Given two intervals of [0, 1] $I = [I_L, I_U] \subseteq [0,1]$ and $J = [J_L, J_U]$, the minimum of both intervals is an interval $K = \mathrm{MIN}(I,J) = [\mathrm{MIN}(I_L, J_L), \mathrm{MIN}(I_U, J_U)]$. Given two evidence sets *A* and *B* defined for each *x* of *X* by:

$$A(x) = \left\{\left\langle I_i^x, m_A^x(I_i)\right\rangle\right\}, \quad i = 1, \ldots, n \qquad \textbf{(41)}$$

and

$$B(x) = \left\{\left\langle J_j^x, m_B^x(J_j)\right\rangle\right\}, \quad j = 1, \ldots, m \qquad \textbf{(42)}$$

where I_i and J_j are intervals of [0,1]. Their intersection is an evidence set $C(x) = A(x) \cap B(x)$, whose intervals of membership K_k and respective basic probability assignment $m_C(K_k)$ are defined by:

$$m_C^x(K_k^x) = \sum_{MIN(I_i^x, J_j^x) = K_k^x} m_A^x(I_i^x) \cdot m_B^x(J_j^x) \qquad \textbf{(43)}$$

6.1.3 Union

The union of two IVFS [Gorzałczany, 1987] is defined as the maximum of their respective lower and upper bounds of their membership intervals. Given two intervals of [0, 1] $I = [I_L, I_U] \subseteq [0,1]$ and $J = [J_L, J_U]$, the maximum of both intervals is an interval $K = \mathrm{MAX}(I,J) = [\mathrm{MAX}(I_L, J_L), \mathrm{MAX}(I_U, J_U)]$. Given two evidence sets *A* and *B* defined by (41) and (42), their union is an evidence set $C(x) = A(x) \cup B(x)$, whose intervals of membership K_k and respective basic probability assignment $m_C(K_k)$ are defined by:

$$m_C^x(K_k^x) = \sum_{MAX(I_i^x, J_j^x) = K_k^x} m_A^x(I_i^x) \cdot m_B^x(J_j^x) \qquad \textbf{(44)}$$

6.1.4 Increasing Uncertainty

By utilizing the connectives (43) and (44), the uncertainty of our models tends to increase, as two bodies of evidence on the unit interval are combined into a new one, by preserving most perspectives (contexts) involved. There will be at least as many intervals in the combined set as the minimum of intervals in the combining sets. In other words, if $|i^x|$ and $|j^x|$ represent the number of intervals (perspectives) present, respectively, in combining sets *A* and *B* for element *x*, then the combined set *C* will have at least MIN($|i^x|,|j^x|$) intervals for concept *x*. An alternative to this way of combining evidence sets is described below.

6.2 Uncertainty Decreasing Operation Between Evidence Sets

We can combine evidence sets by preserving all their perspectives (though with reduced weights as the joined basic assignment must still add up to 1) as above, thus increasing the uncertainty complexity, or we can combine them only according to the coherent perspectives (those aiming at the same intervals) by utilizing Dempster's rule of combination (7) presented in section 2.24 , and decrease the uncertainty complexity. Given two evidence sets *A* and *B* defined by (41) and (24), their uncertainty decreasing combination is an evidence set $C(x) = A(x) \otimes B(x)$, whose intervals of membership K_k and respective basic probability assignment $m_C(K_k)$ are defined by:

$$m_C^x(K_k^x) = \frac{\displaystyle\sum_{I_i^x \cap J_j^x = K_k^x} m_A^x(I_i^x) \cdot m_B^x(J_j^x)}{1 - \mathbf{K}}$$
$$\text{for } K_k^x \neq \varnothing, \text{ where } \mathbf{K} = \sum_{I_i^x \cap J_j^x = \varnothing} m_A^x(I_i^x) \cdot m_B^x(J_j^x) \qquad \textbf{(45)}$$

This operation eliminates all focal elements which do not coincide (or intersect) in both bodies of evidence being combined, while the operations of section 6.1 maintain some evidential weight for these, though enhancing those that do intersect.

Dempster's rule of combination is used to combine different bodies of evidence over the same frame of discernment. It is an all or nothing rule, that is, if the focal elements of two distinct bodies of evidence being combined are disjoint, no combination is possible. In this situation, in DST, if we still consider that there is relevant interaction between the two bodies of evidence which our frame of discernment cannot capture, then we either rethink our basic probability assignments or the frame of discernment is changed by introducing new primitives common to both bodies of evidence. Now consider that our model of categorization, by utilizing Dempster's rule, reaches a combination of categories whose bodies of evidence are completely incoherent. That is, no new category is obtainable. If this result is reached in some intermediate step of an approximate reasoning process, the process is naturally stopped. To be able to continue with this process, we have to obtain some transitional category. Since the frame of discernment of the belief attributes of an evidence set is the unit interval, we cannot aim to refine it in any way. For this reason, I have proposed uncertainty decreasing and increasing operations for evidence sets. If the evidence sets being combined are at least partially coherent, we can use Dempster's rule which will reduce the uncertainty present. If this coherency is not attainable, we can choose an uncertainty increasing operation which largely maintains the evidence from both structures being combined, until a more coherent state of evidence is encountered at a later stage.

The uncertainty decreasing operation can be used when we have coherent evidence of membership in combining evidence sets, and when we wish to reduce dramatically the amount of uncertainty present in some simulation of human reasoning processes. In an artificial system, this operation might be identified we fast decision-making processes. Say, if we possess two categories which must be combined in order to make a fast decision, then uncertainty must be reduced and the most coherent result chosen. On the other hand, if we do not have coherent membership evidence, or if we do not need to engage in fast decision making, but instead desire to search for more conflictuous, far-fetched, associations (from wildly different contexts), then the uncertainty increasing operations should be chosen.

6.3 The Pet-Fish Example

One of the traditional examples of complex category formation used to illustrate the problems of the fuzzy logic approach, is the formation of a *Pet-Fish* Category from constituent categories *Pet* and *Fish*. The category of *Pet-Fish* is obtained here by traditional intersection (minimum operator) of constituent fuzzy sets. As we can see, the problem is that we wanted a higher value for *Guppy* in the *Pet-Fish* category than in either of its constituents. We could have obtained that value by utilizing another operator such as a 'compensatory-AND' [Zimmerman and Zysno, 1980]; however, that would also cause the membership values of the other elements to increase to undesirable results (figure 5). The bottom line is that with a pure fuzzy set model, each element of the categories being combined needs a different aggregation operation — this makes fuzzy set models very cumbersome.

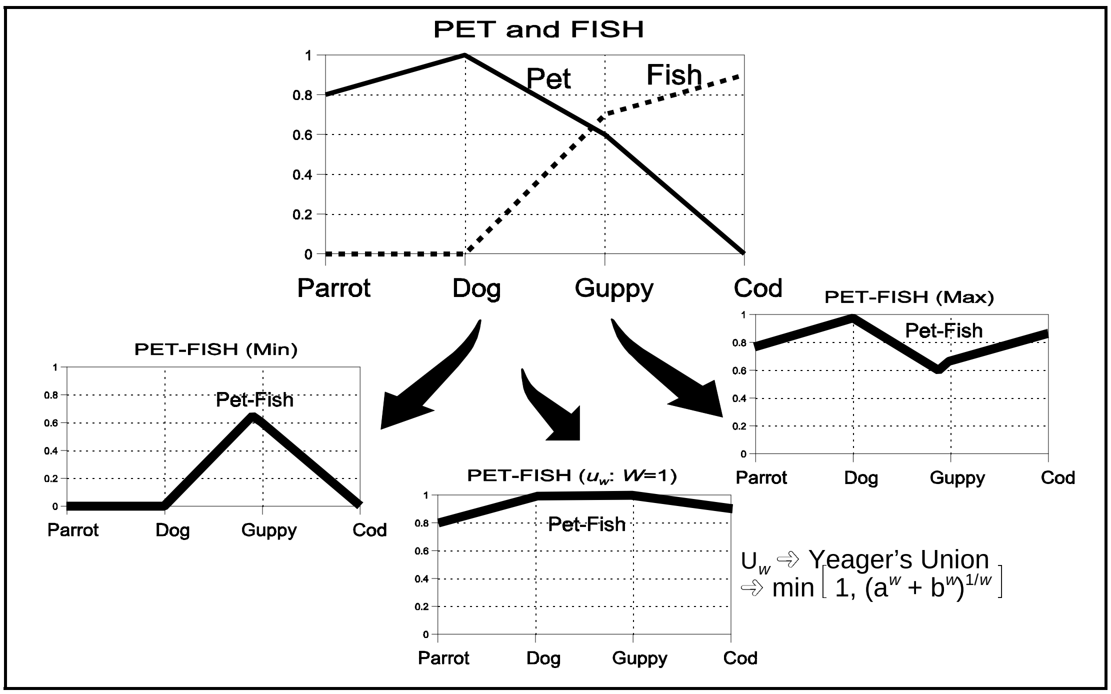

Figure 5: Fuzzy Set model of Pet-Fish. No single connective will yield a satisfying result for the whole category.

According to Türkşen's [1986] construction of IVFS's based on normal forms, the intersection of two fuzzy sets, *A* and *B*, results in an IVFS *I*:

$$I(A \cap B) = \{ I \,|\, A \cap B \subseteq I \subseteq (A \cup B) \cap (A \cup B^{c}) \cap (A^{c} \cup B) \}$$

I have used here both the traditional max-min operators as well as the conjugate pair of t-norms and t-conorms referred to as bold intersection and union: $T(a, b) = \max(0, a + b - 1)$ and $S(a, b) = \min(1,$

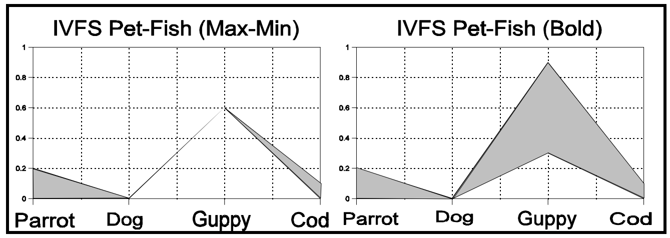

Figure 6: IVFS Model of Pet-Fish, with max-min and bold T-norm T-conorm pairs

a + *b*), for the same constituent categories *Pet* and *Fish* of the previous example. We can see that in this particular case the max-min operators did not yield very good results, though the bold operators give a good representation of what we were looking for (figure 6). Notice, particularly in the bold case, how the IVFS representation enhances and accounts for the char acteristics of **<u>both</u>** constituent concepts. In the case of *Guppy*, the result is a large interval of membership (high nonspecificity), which simply represents our lack of knowledge regarding this element: its membership can vary within a large interval which acknowledges the need for context information; something the fuzzy set model could not accomplish. Also, in this case, one single operation was enough to obtain coherent results for the entire category. However, context is not explicitly accounted for, also contextual conflict is not allowed. Another shortcoming refers to the fact that this model acknowledges the increase in uncertainty when two fuzzy sets are combined (into an IVFS), but how should two IVFS themselves be combined? Should the uncertainty type increase again?

Figure 7 shows how evidence sets can be used to better model this situation. Now, the category of Pet allows for three different contexts: “playful, Interactive, Companion”, “Controllable and Tidy”, and “Cuddly and Cute”. I have graded these contexts, respectively, with the following values of the basic probability assignment: {0.5, 0.1, 0.4}. Fish is given by a simple IVFS. Notice that Guppy in the category of Pet instead of being ascribed a very fuzzy value of membership as the fuzzy set representation did (dotted line in figure 7), receives instead a very conflicting, contextual, not very fuzzy, membership. This shows that actually our positioning of Guppy in the category of Pet is very contextually dependent. Figure 7 exemplifies what the several operations yield for this combination. Notice that the uncertainty decreasing operation, yields a very accurate category of Pet-Fish for this toy example.

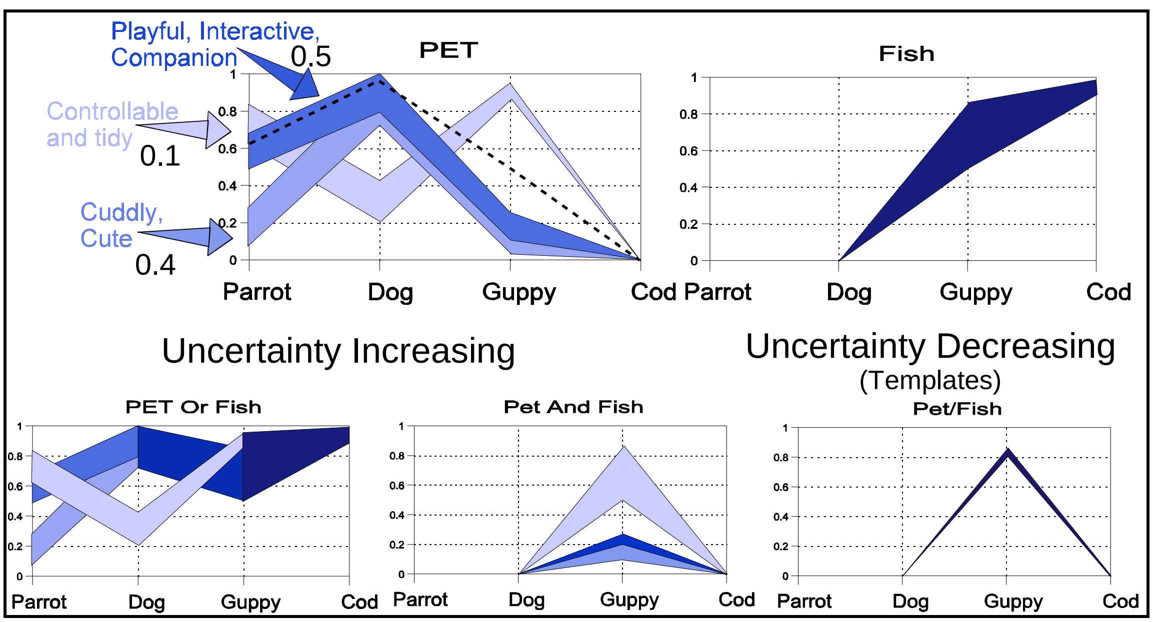

Figure 7: Example of the Belief-Constrained Approximate Reasoning Operations in the Pet-Fish Example

7. Evidence Sets and Evidence Theory

So far, I have discussed set structures as models of cognitive categories, from crisp sets to evidence sets I have stressed that any mathematical model of cognitive categories must offer (i) degrees of inclusion in the category/set, (ii) an accurate account of uncertainty forms in their membership values, and (iii) a way in for context-dependencies and subjective aspects of categories. I have proposed that evidence sets fulfill these three requirements. A natural question now is, why is Evidence Theory not enough by itself to effectively model cognitive categories?

Evidence theory is usually thought of in terms of universes of possibilities, that is, frames of discernment. A subset of such an universe is understood as representing the possible values for some proposition, "thus the propositions of interest are in a one-to-one correspondence with the subsets of [the frame of discernment]" [Shafer, 1976, page 36]. Alternatively, we can also think of the frame of discernment as the universe of possible values for a variable x. If our variable represents the possible elements of a universe of discourse, then a category can be defined as a body of evidence defined on such universe. Each focal element, can be seen as a possible perspective for the category.

7.1 Upper and Lower Probabilities Interpretation

Let us consider that a category is defined by a body of evidence $(\mathfrak{F}, m)$ on a universal set X. In other words, the category will be defined by a set $\mathfrak{F}$ of subsets of X (focal elements) with associated basic probability assignment m. Plausibility and belief measures can be constructed from $(\mathfrak{F}, m)$ as defined in section 2.2. Following Dempster's [1967] original interpretation of plausibility and belief measures as upper and lower probabilities, respectively, we can understand these probability limits as offering a nonspecific (interval-valued) membership of subsets of X in the category, which would satisfy the first requirement above. Nonetheless, several problems are encountered with this model of categories. First, notice that the basic probability assignment values must add up to one (eq. 3), this constrains the category as it introduces a dependency on its elements. That is, because of the probabilistic constraint, the value of membership of an element, which would be given here by the belief-plausibility interval, would be constrained by the value of membership of other elements. Specifically, their individual membership is not free to attain any value as it is desired of a set structure or a cognitive category.

Furthermore, membership in a category (upper and lower probabilities) is not attributed to singletons but to subsets of the universal set. In addition, the second and third requirements would not be satisfied as conflict would not be captured in the individual membership values and no account of context is included.

7.2 Belief Interpretation

Consider now that a category is still defined by the body of evidence $(\mathfrak{F}, m)$, only now, more in line with Shafer's [1976] interpretation, the basic probability assignment function m will identify the portions of belief ascribed exactly to the focal elements $\mathfrak{F}$. This way, each exact portion of belief and its associated focal element can be related to a particular context in a larger imbedding model. In other words, the sort of categories we obtain with this interpretation are formed by crisp subsets of the frame of discernment with associated belief values: membership is all or nothing, but belief is graded. In a way, we have classic categories with an account of belief, subjectivity, and a way in for context-dependencies in a larger model of categorization. Clearly, this interpretation satisfies the third requirement but not the first and the second (remember that accurate accounts of uncertainty forms are desired in membership values, not in the definitions of focal elements in the universal set *X*).

7.3 Generalized Dempster-Shafer Theory

Several ways of extending the Dempster-Shafer theory of evidence into a fuzzy set framework have been proposed, but probably the most general and well known approach is John Yen's [1990] generalization. Basically, though with many interesting consequences beyond the scope of this dissertation, the idea is to move from crisp focal elements to fuzzy focal elements. In this case, we no longer have classical, all-or-none, categories but introduce degrees of membership, thus satisfying the first requirement in addition to the third requirement already satisfied by the second interpretation of evidence theory in the previous section. Naturally, to satisfy the second requirement, that is, to obtain an accurate account of uncertainty forms in the membership degrees of a set/category's elements, we can extend the fuzzy focal elements to interval-valued focal elements, or even more generally to sets of fuzzy sets. This seems to satisfy all of the three requirements above, so, why are evidence sets preferable over generalized evidence theory as models of categories? The next subsection should answer this question.

7.4 Evidence Sets : Independent Membership

Evidence sets have unconstrained membership; that is, the values of membership for each element *x* (singleton) of the universal set *X* are independent of each other. In contrast, the categories defined solely with evidence theory in the previous sections, are set oriented, that is, they define categories with focal elements which are subsets of *X*. Thus, the evidence a particular context offers is associated with a set of singletons rather than with a singleton itself. Naturally, a singleton can also be represented by a set, but if focal elements are singletons, then we will need many focal elements to represent a category, and since their respective evidential weights given by the basic probability assignment must add up to one, each singleton will necessarily have a small degree of belief associated with it. In other words, the belief we have that a certain singleton belongs to a category, will be dependent on the belief we ascribe to other singletons. This kind of dependence is not desirable of a model of a category. Like a set structure, the inclusion of an element in a category should not necessarily be dependent on other elements already included in it. A larger model of categorization may impose these constraints at a higher level, but the basic mathematical structures used should not impose them at the onset.

An evidence set allows a complete separation of membership and belief between elements in a category since an account of belief is not used to constrain the elements of the universal set but to constrain their <u>individual</u> membership values in the unit interval. Thus, the membership/belief of an element *x* is independent from that of another element *y*. It is important to realize that belief is still constrained <u>for each individual</u> membership qualification, in other words, the basic probability assignment used to qualify the possible intervals of membership, must still add up to one. This is so because even though the membership of an element in a category should not be constrained by that of another (free membership), the evidential qualification of possible intervals of membership for each element must add to one to maintain Shafer's [1976] convention that the total belief in an evidential situation has measure one (constrained belief) — all recognizable evidence, regarding a particular element, is made to add up to one. With this independent quantifiability of membership/belief for each element in a universal set, the contexts that affect an element's membership in a category can be completely different from element to element, a desirable characteristic for radial categories. A graphical comparison of categories as modeled by DST, Generalized DST, and evidence sets is shown in figure 8.

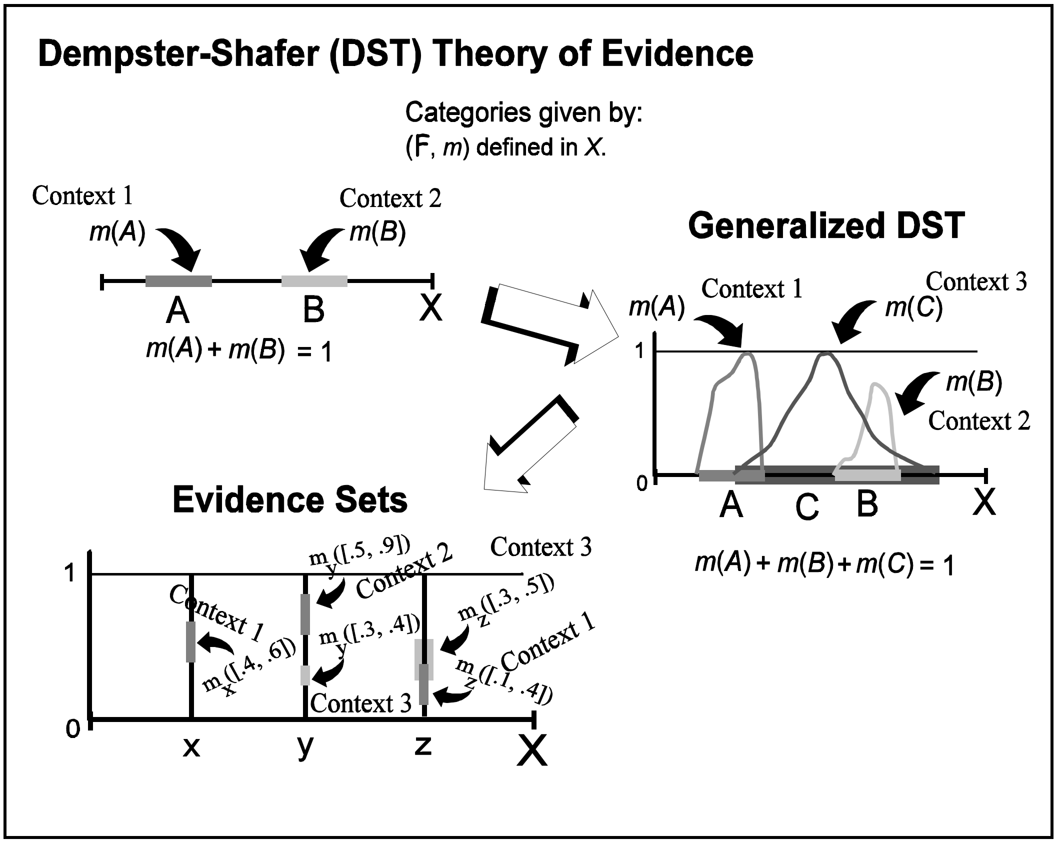

Figure 1: DST captures categories as all-or-nothing sets which are subjectively, and contextually constrained. Generalized DST adds fuzzy membership, but still membership is not independent but probabilistically constrained. Evidence sets provide independent membership and a constrained, belief-based qualification of this membership.