Uncertainty Measures for Evidential Reasoning I: A Review

Nikhil R. Pal and James C. Bezdek
Division of Computer Science, University of West Florida, Pensacola, Florida

Rohan Hemasinha
Department of Mathematics and Statistics, University of West Florida, Pensacola, Florida

ABSTRACT

This paper is divided into two parts. Part I discusses limitations of the measures of global uncertainty of Lamata and Moral and total uncertainty of Klir and Ramer. We prove several properties of different nonspecificity measures. The computational complexity of different total uncertainty measures is discussed. The need for a new measure of total uncertainty is established in Part I. In Part II, we propose a set of intuitively desirable axioms for a measure of total uncertainty and then derive an expression for the same. Several theorems are proved about the new measure. The proposed measure is additive, and unlike other measures, has a unique maximum. This new measure reduces to Shannon's probabilistic entropy when the basic probability assignment focuses only on singletons. On the other hand, complete ignorance—basic assignment focusing only on the entire set, as a whole—reduces it to Hartley's measure of information. The computational complexity of the proposed measure is $O(N)$, whereas the previous measures are $O(N^2)$.

KEYWORDS: ***conflict, confusion, evidential reasoning, entropy, dissonance, specificity, uncertainty***

1. INTRODUCTION

Consider a simple experiment with a six-faced die. Suppose the die is thoroughly shaken and placed on a table covered with a box and you are

---

Address correspondence to James C. Bezdek, Division of Computer Science, University of West Florida, Pensacola, FL 32514. N. R. Pal is on leave from ISI, Calcutta, India.

Received December 1, 1991; accepted May 1, 1992.

This research was supported by NSF IRI-9003252.

International Journal of Approximate Reasoning 1992; 7:165–183

655 Avenue of the Americas, New York, NY 10010 0888-613X/92/$5.00

asked to guess the top face of the die. To answer this question one faces a type of uncertainty that can be attributed to *randomness* present in the system (experiment). The best answer to this question might be to describe the status of the die in terms of a probability distribution over the different faces (if known). Uncertainty that arises due to randomness in the system is called *probabilistic uncertainty*.

To make the system more complex, suppose an artificial vision system analyzes a digital image of the die and, based on the evidence gathered, suggests that the top face is 5 or 6 with a belief value (confidence) of, say, 0.8. In other words, based on the evidence the system is not able to specify the top face of the die exactly. This kind of uncertainty usually arises due to limitations of the evidence gathering and interpretation system. Uncertainty in this second situation is due to a difficulty in specifying the exact solution and is called *nonspecificity* in the literature.

Finally, suppose you are asked to interpret the top most face of the die as, *high* or *low*. In this case a different type of uncertainty (ambiguity) is faced. Here the existence of a fuzzy event *high* is assumed and one has to gauge the extent to which this event has occurred. In a probabilistic experiment, an event either occurs or not, but here an event may occur partially. This type of uncertainty arises due to the presence of *fuzziness* in the system.

It is clear that fuzzy uncertainty differs from probability and nonspecificity. Fuzzy uncertainty deals with situations where boundaries of the sets under consideration are not sharply defined—partial occurrence of an event. On the other hand, for probabilistic and nonspecific uncertainties there is no ambiguity about set–boundaries, but rather, about the belongingness of elements or events to crisp sets. The present study confines itself only to non-fuzzy uncertainties. The literature is quite rich on fuzzy uncertainty; interested readers may refer to [1–4].

Intuitively one feels that uncertainty due to nonspecificity is related to probabilistic uncertainty. It is very difficult to look at either of them in isolation. For example, when belief values are assigned only to singleton elements, there is no uncertainty due to nonspecificity, only randomness may be present. But if a belief value is assigned to a set with cardinality more than one, then there is only nonspecificity. On the other hand, when belief values are assigned to more than one set that are not singletons, one faces both randomness and nonspecificity.

In our discussion, we restrict ourselves to a finite reference set X, and an unknown element x belonging to X. It is also assumed that all information about the belongingness of x to X, if available, is expressible in the language of the Dempster–Shafer [5] theory of evidence. Several authors have suggested different measures for uncertainty. Yager [6] proposed a measure called dissonance or conflict, while Hohle [7, 8] suggested a measure to quantify the level of confusion present in a body of evidence. Smets [9] suggested a different type of measure for the information content of an evidence. Unlike the measures of Yager [6] and Hohle [7, 8], the measure of Smets is not a generalization of Shannon's entropy. Higashi and Klir [10] proposed a measures of nonspecificity for a possibility distribution that was later extended to any body of evidence by Dubois and Prade [11]. Recently Klir and Ramer [12] pointed out some limitations of the measure of conflict (confusion) of Hohle [7, 8], and suggested a new measure for the same. We give an example which shows that this measure also leads to an unappealing situation, and we prove several theorems on different nonspecificity measures.

Lamata and Moral [13] proposed two *composite* measures, called *global* uncertainty measures, that attempt to quantify both the probabilistic and nonspecific aspects of uncertainty. One of their measures simply adds the dissonance measure of Yager to the nonspecificity measure of Dubois and Prade; whereas the other one is introduced via definition without prior motivation or justification. Klir and Ramer subsequently suggested another composite measure called *total* uncertainty that is defined as the sum of Dubois and Prade's nonspecificity and a new measure of conflict called discord [12]. Composite measures reflect some interesting aspects of uncertainty, but analysis reveals that they can lead to intuitively unappealing situations when interpreted as *total* uncertainty. For example, all of these composite measures have several maxima, which makes it difficult to gauge the quality of evidence based on their numerical values. Moreover, the computational overhead for each of these measures is high. Finally, and most importantly, elementary measures of nonspecificity or probabilistic uncertainty such as dissonance, discord, or nonspecificity attempt to measure only one of the two aspects of non-fuzzy uncertainty, so interpretation of a composite measure such as the *total* or *global* uncertainty is difficult. Aggregation of elementary measures may depend on the mode of interaction between different aspects of uncertainty represented by the components in the sum. Because nonspecificity and randomness are related in an unknown way, it does not seem desirable to add expressions for these measures directly to get a measure for total uncertainty.

In Part I of this paper we evaluate existing measures of uncertainty and prove several new properties about some of them. We show that existing measures cannot model situations when *no* evidence is better than *inconsistent* evidence. To elucidate this point, consider the following two situations: First, an expert is completely ignorant about the unknown element (because of insufficient or no evidence), and the confidence (belief) assigned to the universal set X is *1*; and second, based on the evidence, an expert assigns a belief value of $1/(2^n - 1)$ to each of the possible nonempty subsets of X. In the former case, the expert is confident about his or her ignorance; whereas in the latter case, the expert seems confused. Because the basic assignment in the second case is expected to be inconsistent, the total uncertainty in the second case should be larger than in the first case. However, none of the existing measures support this position.

All of the problems itemized above motivate us to look for a new measure of average total uncertainty. Our approach will be to postulate a set of axioms that seem desirable for any measure of total uncertainty—as opposed to axioms related to only one component in a composite sum—and then to derive a function that satisfies the axioms. The measure of total uncertainty we discover will account for the expert's dilemma given above. Several theoretical properties of the new measure are studied, and a numerical example is given that affords an empirical comparison with previous measures. Unlike other measures of total uncertainty, the new measure has a unique maximum. Shannon's probabilistic entropy and Hartley's entropy are shown to be special cases of our measure. Finally, we show that the new measure is computationally more tractable than previous measures.

2. BASIC TERMINOLOGY

In this section we introduce the basic terminology and definitions of the Dempster-Shafer theory of evidence. Let X be a finite universe of discourse, $|X| = n$, $P(X)$ the power set of X, and x any element in X. All information about the belongingness of x to X is expressible by a basic probability assignment (BPA) function m: $P(X) \to [0, 1]$ that satisfies:

$$m(\phi) = 0 \quad (\phi = \text{empty set}) \tag{1}$$

$$\text{and} \quad \sum_{A \subseteq X} m(A) = 1. \tag{2}$$

The value $m(A)$ represents the degree of evidence or belief that the element x in question belongs exactly to the set A but not to any B such that $B \subset A$. The pair (F, m) is called the *body of evidence* for x, where F is the set of all subsets A of X such that $m(A) > 0$. Elements of F are called *focal elements*. If the focal elements are nested (i.e., can be arranged in a sequence such as $A_1 \subset A_2 \subset \cdots \subset A_k, \ldots$) then the corresponding body of evidence is called a *consonant* body of evidence. In this context the following observations about a body of evidence may be made. If $|F| = 1$ and $A \in F$ then either $|A| = 1$, and there is no uncertainty; or $|A| > 1$, and there is uncertainty due to nonspecificity. Conversely, $|F| > 1$ and $|A| = 1$ for all $A \in F$ represents a situation with only randomness. In all other cases both randomness and nonspecificity will be present, because when $|F| > 1$ and $|A| > 1$, at least for some $A \in F$, the element in question can be in any one of the sets (focal elements) and given the focal set, it can be any member of the set.

Shafer [5] defined two fuzzy measures on a body of evidence (F, m), namely, *Belief* (Bel) and *Plausibility* (Pl) as follows:

$$Bel(A) = \sum_{B \subseteq A \in F} m(B); \tag{3}$$

$$\text{and} \quad Pl(A) = 1 - Bel(A^c) = \sum_{A \cap B \neq \phi} m(B) \tag{4}$$

where A^c is the complement of A. Shafer also defined the *Commonality Number* of A as follows:

$$Cm(A) = \sum_{A \subseteq B} m(B) \tag{5}$$

Conceptually $Bel(A)$ represents the total degree of evidence that the concerned element belongs to A and/or some of its subsets. On the other hand, $Pl(A)$ not only gives the total degree of evidence that the concerned element belongs to A or some of its subsets, but also to those sets that have nonempty intersections with A. Properties of these two measures can be found in [1, 5, 14]. A belief function (Bel) is called a *vacuous* belief function if $Bel(X) = 1$ and $Bel(A) = 0$ for $A \neq X$. For a vacuous belief function $m(X) = 1$ and $m(A) = 0$ for $A \neq X$. For a consonant body of evidence the belief and plausibility measures are called *necessity* and *possibility* measures, respectively. The commonality function $Cm(A)$ gathers pieces of evidence supported by A. The commonality function plays the role of belief for a conjunctive body of evidence [14]. Any of the four set functions m, Bel, Pl and Cm can be expressed uniquely in terms of any other [5].

Lastly, we note that every possibility measure π on $P(X)$ is uniquely determined by a possibility distribution function r: $X \to [0, 1]$ via the formula:

$$\pi(A) = \max_{x \in A}\{r(x)\}$$

A possibility distribution $r = (\rho_1, \rho_2, \ldots, \rho_n)$, $\rho_i = r(x_i)$ is called an *ordered possibility distribution* if $\rho_i \geq \rho_j$ when $i < j$. For a possibility measure we can assume without any loss of generality that the focal elements are some or all of the subsets in the complete sequence of nested subsets:

$$A_1 \subset A_2 \subset \cdots \subset A_n \ (= X) \quad \text{where } A_i = \{x_1, \ldots, x_i\},\ i = 1, \ldots, n$$

Note that $m(A) = 0$ for each $A \neq A_i$, and it is not required that $m(A_i) \neq 0, \forall\ i = 1, \ldots, n$. In this context there is a one to one correspondence between the possibility distribution and *BPA* [1]:

$$m(A_i) = \rho_i - \rho_{i+1};\ \rho_{n+1} = 0$$

$$\text{and} \quad \rho_i = \sum_{k=i}^{n} m(A_k)$$

3. MEASURES OF UNCERTAINTY

This section reviews some extant measures for each of the two different aspects of non-fuzzy uncertainty, namely, nonspecificity and conflict/confusion/discord. Table 1 summarizes different existing measures.

The word "probabilistic" in column three of Table 1 does not necessarily indicate a function of a probability distribution. It is used to indicate that part of the overall uncertainty which is due to randomness or chance. Let M_n denote the set of all basic assignments on the power set with 2^n elements. A measures of uncertainty is a mapping S: $M_n \to [0, \infty)$ that satisfies some intuitive notion of uncertainty. In what follows we use *Log* to specify logarithms to some base $a > 1$. Usually, $a = 2$ or $a = e$; different writers have used different bases. However, because a change of base amounts to a multiplicative constant, we omit the base unless clarity demands it.

Yager [6, 15] proposed the following measure of dissonance (conflict) in a body of evidence:

$$E(M) = -\sum_{A \in F} m(A) Log Pl(A). \tag{6}$$

This measure is based on the following interpretation of conflicting evidence: whenever the evidence suggests that the element of concern may belong to either of two or more disjoint subsets, then we have conflicting evidence. In [6, 15] Yager introduced the concept of specificity associated with a possibility distribution. For a normalized possibility distribution r specificity is defined as

$$S(r) = \int_0^1 \frac{1}{|r_\alpha|}\, d\alpha \tag{7a}$$

where $r_\alpha = \{x \mid r(x) \geq \alpha,\ x \in X\}$ and $|*|$ denotes set cardinality. $S(r)$ estimates the precision of a fuzzy set—in other words, it quantifies the extent to which a fuzzy set restricts a small number of values for a variable. A probabilistic interpretation of S is given in [11]. This measure of specificity was subsequently extended (generalized) by Yager [6] to any belief structure as follows:

$$N(m) = \sum_{A \in F} \{m(A)/|A|\}. \tag{7b}$$

$J(m) = 1 - N(m)$ is viewed as a measure of *nonspecificity*. Properties of these measures have been investigated by several authors and can be found in[6, 13, 15, 16].

Hohle [7, 8] proposed the following measure of *confusion*:

$$C(m) = -\sum_{A \in F} m(A) LogBel(A) \tag{8}$$

C purports to represent the conflict that arises when two evidential claims $m(A)$ and $m(B)$ conflict within the same body of evidence and when $B \not\subset A$. Properties of E, N and C have been studied by Dubois and Prade [16].

Higashi and Klir [10] derived a measure of nonspecificity (called U-uncertainty) for possibility distributions based on a set of desirable axioms. U-uncertainty was proposed to be the possibilistic counter part of Shannon's entropy and at the same time, a generalization of the Hartley information measure. U-uncertainty is thus derived so as to satisfy all properties of Shannon's entropy for which possibilistic counterparts are meaningful, as well as to generalize Hartley's information. U-uncertainty for an ordered possibility distribution [1] is given by:

$$U(r) = \sum_{i=1}^{n} (\rho_i - \rho_{i+1}) Log|A_i| \tag{9}$$

Recall that $A_i = \{x_1, x_2, \ldots, x_i\}$; $\rho_i = r(x_i)$, r is the possibility distribution function and $\rho_{n+1} = 0$. Since $\rho_i - \rho_{i+1} = m(A_i)$, where m is the *BPA* associated with the possibility distribution, the expression for $U(r)$ can be rewritten as;

$$U(r) = \sum_{i=1}^{n} m(A_i) Log|A_i|. \tag{10}$$

Dubois and Prade [11] made a natural extension of this measure for any *BPA* as follows:

$$I(m) = \sum_{A \in F} m(A) Log|A|. \tag{11}$$

Properties of I were studied by Lamata and Moral [13], who suggested two *composite* measures for *global uncertainty*: The first one was defined as the sum of the measures of conflict (E) nonspecificity (I):

$$G_1(m) = E(m) + I(m). \tag{12}$$

Lamata and Moral [13] point out that measures like (11) are difficult to extend (or generalize) to a more general class of fuzzy measures because such a measure is defined in terms of a BPA, and, for example, in the case of ordered fuzzy measures there is no function similar to the BPA. We add here that extension of measures such as I or E to any fuzzy measure may in fact be entirely disconnected from the idea of uncertainty assessment in the context of evidential reasoning. To circumvent this problem Lamata and Moral suggested a second expression for global uncertainty:

$$G_2(m) = EV_{Bel}(f), \quad \text{where for any } a \in X, \tag{13}$$

$$f(a) = -Log\left(\frac{Pl(\{a\})}{\sum_{b \in X} Pl(\{b\})}\right); \tag{14a}$$

$$\text{and} \quad EV_{Bel}(f) = \int_0^\infty Bel\{x \in X/f(x) \geq \alpha\}\, d\alpha \tag{14b}$$

is the expected value of f with respect to *Bel*—in fact, however, G_2 does *not* circumvent the problem identified by Lamata and Moral, because it is also defined in terms of a *BPA*. The expression for $G_2(m)$ can be decomposed and rewritten as

$$G_2(m) = V(m) + W(m), \tag{15}$$

$$\text{where} \quad V(m) = EV_{Bel}(-Log(Pl(\{x\}))); \tag{16}$$

$$\text{and} \quad W(m) = Log\left(\sum_{A \subseteq X} m(A)|A|\right) \tag{17}$$

Lamata and Moral interpreted $-Log\,Pl(\{x\})$ as the degree of *surprise* experienced when it is known that x is the unknown element belonging to X. $EV_{Bel}(-Log\,Pl(\{x\}))$ is the expected value of surprise with respect to *Bel*. They considered V the *lowest* degree of surprise because the expectation is computed with respect to *Bel*. The second term W is viewed as a measure of *imprecision*. In fact, $\Sigma m(A)|A|$ is the average *cardinality* of the focal elements. Hence the second term can be interpreted as a measure of nonspecificity. In contrast to the average cardinality of focal elements, Yager's specificity measure (7b) gives an indication of the *dispersion* of belief. In Lamata and Moral [13] no motivation is given for defining G_2 as in (15). Consequently, it is hard to us to offer an explanation for the interpretation of the "lowest expected degree of surprise."

According to Klir and Ramer the measure of dissonance at (6) is unsatisfactory; they feel that $m(B)$ conflicts with $m(A)$ whenever $B \not\subset A$, not only when $B \cap A = \varnothing$ [12]. Hohle's measure of confusion accounts for this problem. However, Klir and Ramer suggest in [12] that the degree of violation of $B \subseteq A$ should influence the value of the conflict measure. With a view towards incorporating this constraint they suggested an alternative measure of conflict [12]:

$$D(m) = -\sum_{A \in F} m(A) Log[1 - Con(A)], \tag{18}$$

$$\text{where} \quad Con(A) = \sum_{B \in F} [m(B)|B - A|/|B|] \tag{19}$$

Klir and Ramer suggest that $Con(A)$ expresses the sum of individual conflicts of evidential claims with respect to a particular set A, each of which is properly scaled by the degree to which the containment $B \subseteq A$ is violated. Equation (18) can be simplified to the following form [12]:

$$D(m) = -\sum_{A \in F} m(A) Log\left[\sum_{B \in F} m(B)|A \cap B|/|B|\right]. \tag{20}$$

Klir and Ramer then defined *total uncertainty* as:

$$T(m) = D(m) + I(m). \tag{21}$$

Comparing (21) and (12), we see that I is common to both G_1 of Lamata and Moral and T of Klir and Ramer. Indeed, the number $D - E$ (D is always $\geq E$) is in some sense a measure of the distinction drawn by these authors between global and total uncertainty.

Measures like E, C, N, D and I etc. (the first seven rows of Table 1) can be called *elementary* measures, whereas functions such as G_1, G_2 and T may be viewed as *composite* measures (the last three rows of Table 1). Composite measures exhibit a trade-off between the assessment objectives of their elementary factors. For example, G_1 balances nonspecificity against dissonance. As long as aspects of uncertainty that are quantified by elementary factors are conceptually additive in nature, summation to form a composite measure is meaningful. Otherwise, interpretation of a composite measure is at best difficult. In view of this, a better approach to the assessment of total uncertainty might be to derive a measure based on a set of axioms (desirable properties) about what it is supposed to quantify. This is the path we follow in part II of this paper.

Table 1. A Catalog of Uncertainty Measures

Author(s)	Sum	Probabilistic	Nonspecific
Yager [6]		$E(m) = -\sum_{A \in F} m(A) Log Pl(A)$ (Dissonance)	$J(m) = 1 - \sum_{A \in F} \{m(A)/\|A\|\}$ (Nonspecifity)
Hohle [7, 8]		$C(m) = -\sum_{A \in F} m(A) Log Bel(A)$ (Confusion)	
Higashi & Klir [10]			$U(r) = \sum_{i=1}^{n} m(A_i) Log \|A_i\|$ (Nonspecificity)
Smets [9]		$L(m) = -\sum_{A \subseteq X} Log Cm(A)$	
Dubois & Prade [11]			$I(m) = \sum_{A \in F} m(A) Log \|A\|$ (Nonspecificity)
Klir & Ramer [12]		$D(m) = -\sum_{A \in F} m(A) Log \left[\sum_{B \in F} m(B) \|A \cap B\| / \|B\| \right]$ (Discord)	
Lamata & Moral [13]		$V(m) = EV_{Bel}(-Log(Pl(\{x\})))$ (Innate Contradiction)	$W(m) = Log\left(\sum_{A \subseteq X} m(A) \|A\| \right)$ (Imprecision)
Lamata & Moral [13]	$G_1(m)$	$= E(m) + I(m)$ (global uncertainty)	
Lamata & Moral [13]	$G_2(m)$	$= V(m) + W(m)$ (global uncertainty)	
Klir & Ramer [12]	$T(m)$	$= D(m) + I(m)$ (total uncertainty)	

We offer the following comments on G_1, G_2 and T: If $Con(A)$ represents the sum of individual conflicts of evidential claims with respect to a particular set A, then the average (or weighted average) measure of conflict preferably should be a direct function of $Con(A)$ (as it is in case of Shannon's entropy). The physical significance of the use of $(1 - Con(A))$ in the expression for D is difficult to interpret except as it helps yield an entropy-like expression. Moreover, when there is more than one focal element, there is some uncertainty due to randomness associated with each focal element. T captures this uncertainty only partially. T and G_1 account for only one aspect (conflict) of uncertainty. We illustrate this with an example: Let $A \subset B$, $m(A) > 0$, $m(B) > 0$, and $m(A) + m(B) = 1$. Using (19) one finds that $Con(A) > 0$ but $Con(B) = 0$, and $D(m) = -m(A)Log(1 - Con(A))$. Similarly, $Pl(A) = Pl(B) = 1$, resulting in $E(m) = 0$. Thus E does not capture any uncertainty due to randomness, whereas D does not take into account the ignorance (due to randomness) associated with the basic assignment $m(B)$. G_2 suffers from the same problem. It is, therefore, questionable whether T, G_1 or G_2 should be called total uncertainty. Next we exhibit a counterintuitive property of $Con(A)$.

EXAMPLE 1. Suppose $X = \{1,2,3,4\}$ and the BPA function is defined as follows: $A = \{1,2,3\}$, $B = \{1,2,3,4\}$; $m(A) = 0.4$, $m(B) = 0.6$. For this basic assignment, because there are only two focal elements, $Con(A)$ should represent the conflict of $m(A)$ with $m(B)$, whereas $Con(B)$ should reflect the conflict of $m(B)$ with $m(A)$. By (19) we get $Con(A) = .4*0 + .6*1/4 = .15$ and $Con(B) = .6*0 + .4*0 = 0.0$. Thus although $A \subset B$, $m(A)$ is in conflict with $m(B)$ and the conflict of A with B is different from the conflict of B with A. Let us consider another BPA on the same X as follows: $A = \{1,2\}$, $B = \{2,3,4\}$, $m(A) = 0.2$ and $m(B) = 0.8$. According to equation (19), $Con(A) = 0.53$ and $Con(B) = 0.1$. It seems reasonable to expect that if A is in conflict with B then B should also be in conflict with A to the same extent. In a private communication [19] Klir has acknowledged that $Con(A)$ has defects, and has proposed yet another measure (strife) that seems to correct the first deficiency ($Con(A) > 0$ even when $A \subset B$) by redefining the expression for $Con(A)$ as $Con(A) = \Sigma_{B \in F} m(B)(|A - B|/|A|)$. However, with the new definition one gets $Con(A) = 0.4$ $Con(B) = 0.13$. Thus (19) (hence (20) too) still seems to be an unappealing measure of conflict.

A different type of information measure was proposed by Smets in [9]. His information measure is based on a set of five desirable properties, of which the most important was additivity. Here additivity means that the information content from two distinct, nonconflictual evidences is the sum of the information content of each evidence. Smets' measure takes the form:

$$L(m) = -\sum_{A \subseteq X} Log\,Cm(A). \tag{22}$$

If the belief structure is non-dogmatic ($m(X) > 0$), then $Cm(A) > 0$ for all $A \subseteq X$ and L is well defined; otherwise L is undefined. L is not a generalization of Shannon's entropy; for a Bayesian belief structure it does not reduce to Shannon's probabilistic entropy, rather, it is undefined. Smets did not provide a satisfactory interpretation of L in the context of uncertainty measurement, and because all of the other measures (that deal with randomness) discussed so far represent some kind of average (expected) values, whereas L does not, our subsequent discussion will ignore this measure.

4. SOME NEW PROPERTIES OF NONSPECIFICITY MEASURES

In the following theorems, we assume $|X| = n$. Let $F_k \subseteq F$ such that $F_k = \{A \mid A \in F \text{ and } |A| = k\}$; and let $P_k = \Sigma_{A \in F_k} m(A)$, so by (2), $\Sigma P_k = 1$.

THEOREM 1 *Redistribution of m keeping P_k fixed over F_k does not change the value of nonspecificity measures I, J, and W.*

Proof

$$I(m) = \sum_{A \in F} m(A) Log|A| = \sum_{k=1}^{n} \sum_{A \in F_k} m(A) Log|A|$$

$$= \sum_{k=1}^{n} \sum_{A \in F_k} m(A) Log|k| = \sum_{k=1}^{n} Log(k) \sum_{A \in F_k} m(A)$$

$$= \sum_{k=1}^{n} P_k Log(k)$$

Because the last expression remains fixed on redistribution of m provided the P_k are fixed over, F_k, the theorem follows. Proofs for J and W are similar. ■

EXAMPLE 2. AN IMPLICATION OF THEOREM 1 Suppose $X = \{1, 2, 3, 4\}$ and two *BPA*s m and m_1 are defined as follows:

m:

$$m(\{1\}) = 1/4$$
$$m(\{1,2\}) = 3/4$$

m_1:

$$m_1(\{1\}) = 1/4$$
$$m_1(\{1,2\}) = 3/24$$
$$m_1(\{1,3\}) = 3/24$$
$$m_1(\{1,4\}) = 1/4$$
$$m_1(\{2,3\}) = 1/12$$
$$m_1(\{2,4\}) = 1/12$$
$$m_1(\{3,4\}) = 1/12$$

For both m and m_1, $P_1 = 1/4$ and $P_2 = 3/4$, but $(F, m) \neq (F_1, m_1)$. Theorem 1 asserts that $I(m)$ and $I(m_1)$ are equal (namely, $I(m) = I(m_1) = 0.75$). According to Yager specificity relates to the degree to which the evidence is pointing to a one element realization [6]. In other words, it represents the average mass per element. In this spirit, Theorem 1 and the above observation are quite natural. However, nonspecificity gives an intuitive notion of lack of specification/preciseness (or inability to specify the correct element) that is connected to the uncertainty due to randomness in some unknown fashion. Under this interpretation, the result $I(m) = I(m_1)$ is undesirable.

THEOREM 2 *For focal elements A and B with $|A| \neq |B|$, any reassignment of m preserving $(m(A) + m(B))$ will increase I, J and W if the reassignment makes m more uniformly distributed with respect to the cardinalities of these focal elements.*

Proof Let $|A| = p$, $|B| = q$, where A and B are two focal elements such that $|A| \neq |B|$. Suppose the reassignment m_1 of m is defined as follows:

$$m_1(C) = m(C), \qquad C \in F \text{ and } C \neq A, B$$
$$m_1(A) = m(A) - \delta; \qquad m_1(B) = m(B) + \delta \text{ with } \delta > 0.$$

$$I(m_1) - I(m) = \{m(A) - \delta\} Log p + \{m(B) + \delta\} Log q$$

$$-m(A) Log p - m(B) Log q = \delta \, Log(q/p),$$

so
$$I(m_1) - I(m) = \begin{Bmatrix} > 0 & \text{if } q > p \\ < 0 & \text{if } q < p \end{Bmatrix}$$

Similarly,

$$J(m_1) - J(m) = \delta\left(\frac{1}{p} - \frac{1}{q}\right) = \begin{Bmatrix} > 0 & \text{if } q > p \\ < 0 & \text{if } q < p \end{Bmatrix}$$

To prove the proposition for W, it is enough to observe that logarithm is a monotonic function and $\Sigma m_1(A)|A| > \Sigma m(A)|A|$ for $q > p$ and $\Sigma m_1(A)|A| < \Sigma m(A)|A|$ for $q < p$. ■

THEOREM 3 *If m_1 and m_2 are BPAs on X generating plausibility structures Pl_1 and Pl_2, respectively such that for each $x \in X$, $Pl_1(\{x\}) \leq Pl_2(\{x\})$, then $W(m_1) \leq W(m_2)$.*

Proof

$$Pl_1(\{x\}) \leq Pl_2(\{x\}) \Rightarrow \sum_{x \in X} Pl_1(\{x\}) \leq \sum_{x \in X} Pl_2(\{x\})$$

$$\Rightarrow \sum_{A \subseteq X} m_1(A)|A| \leq \sum_{A \subseteq X} m_2(A)|A|$$

as $\sum_{x \in X} Pl(\{x\}) = \sum_{A \subseteq X} m(A)|A|$ [6,13] ■

THEOREM 4 *W is minimum ($W(m) = 0$) over M_n if and only if m is a Bayesian belief structure (ie, m focuses only on singleton elements).*

Proof Because $\Sigma m(A) = 1$ and $|A| \geq 1$, $W(m) = Log(\Sigma_{A \subseteq X} m(A)|A|) \geq 0$. This shows that the global minimum of W is 0. If m is a Bayesian belief structure then $W(m) = 0$. Conversely, suppose $W(m) = 0$: $W(m) = 0 \Rightarrow \Sigma_{A \in F} m(A)|A| = 1$. Now because $\Sigma m(A) = 1$, if there is any $A \in F$ such that $|A| > 1$, then $\Sigma_{A \in F} m(A)|A| > 1$. Hence $\Sigma_{A \in F} m(A)|A| = 1 \Rightarrow |A| = 1$ for all $A \in F$. ■

THEOREM 5 *W is maximum ($W(m) = Log(n)$) if and only if m is a vacuous belief function.*

Proof W is maximum when $\Sigma_{A \in F} m(A)|A|$ is maximum. Because $\Sigma m(A) = 1$ and the maximum $|A|$ is n, $\Sigma_{A \in F} m(A)|A| \leq n$. If m is vacuous, then $m(X) = 1$ and $\Sigma_{A \in F} m(A)|A| = n$, so W attains the maximum value of $Log(n)$. Conversely, suppose W is maximum, ie, $\Sigma_{A \in F} m(A)|A| = n$. Then $m(X) = 1$, because if $m(X) \neq 1$ then there is at least one $A \in F$, such that $|A| < n$, and hence $\Sigma_{A \in F} m(A)|A| < n$ (contradiction). ■

5. SOME COMMENTS ON COMPOSITE MEASURES T, G_1 AND G_2

This section investigates the extent to which the composite measures T, G_1 and G_2 conform to intuitive notions of *total* uncertainty. We begin with an example.

EXAMPLE 3 Define seven basic probability assignment functions on $X = \{1,2,3,4\}$ as follows:

$\mathbf{m}_1$: $m(\{1\}) = 1.$

$\mathbf{m}_2$: $m(X) = 1.$

$\mathbf{m}_3$: $m(\{1\}) = m(\{2\}) = m(\{3\}) = m(\{4\}) = 1/4.$

$\mathbf{m}_4$: $m(\{1,2\}) = m(\{2,3\}) = m(\{3,4\}) = m(\{1,4\}) = 1/4.$

$\mathbf{m}_5$: $m(\{1,2\}) = m(\{1,3\}) = m(\{1,4\}) = m(\{2,3\}) = m(\{2,4\}) = m(\{3,4\}) = 1/6$

$\mathbf{m}_6$: $m(A) = 1/15$ *for all* $A \in P(X)$, $A \neq \emptyset$.

$\mathbf{m}_7$:

$$
\begin{aligned}
m(A) &= 1/32 \quad \text{if } |A| = 1\\
m(A) &= 2/32 \quad \text{if } |A| = 2\\
m(A) &= 3/32 \quad \text{if } |A| = 3\\
m(A) &= 4/32 \quad \text{if } |A| = 4.
\end{aligned}
$$

Based on our intuitive feeling about *total* uncertainty we offer the following comments. Clearly there is no uncertainty associated with m_1. For m_2, the total uncertainty is due purely to nonspecificity, (randomness is absent). Conversely, m_3 has only uncertainty due to randomness, but not nonspecificity. On the other hand, m_4, m_5, m_6 and m_7 have uncertainties due to both nonspecificity and randomness. It seems plausible to expect the total uncertainty (TU) for these seven *BPA*s to satisfy the inequalities: $TU(m_1) < TU(m_2) = TU(m_3) < TU(m_4) < TU(m_5) < TU(m_6) < TU(m_7)$. $TU(m_2)$ should equal $TU(m_3)$, as m_2 and m_3 are the extreme cases of only randomness or nonspecificity, respectively. m_2 represents a situation when there is only nonspecificity and that to the maximum extent (no randomness); whereas m_3 does not possess uncertainty due to nonspecificity. m_3 has only randomness in a most ambiguous way. The equality of $TU(m_2)$ and $TU(m_3)$ essentially constrains TU to behave "symmetrically" at its extremes. In other words, if there is only one type of uncertainty then the maximum value should be the same in either case.

However, when both types of uncertainty are present, then TU should increase/decrease depending on the complexity of the system (nature of the basic assignment function). In m_4 the amount of uncertainty due to randomness appears to be the same as that of m_3 because in either case (m_3 or m_4) any one of the four possibilities could be true with a confidence value of $1/4$. However, in m_4 there is some uncertainty due to nonspecificity which is absent in the case of m_3. Therefore, $TU(m_3)$ should be strictly less than $TU(m_4)$. Comparison of m_4 and m_5 indicates that m_5 has more randomness than m_4. Even if we assume that nonspecificity is the same for m_4 and m_5, $TU(m_5)$ should be greater than $TU(m_4)$ as m_5 has more focal elements with uniformly distributed belief values. Note that m_3, m_4 and m_5 all distribute belief values (basic assignments) uniformly over their respective set of focal elements. All focal elements in each of the cases have the same cardinality. On the other hand, m_6 distributes the basic assignments uniformly over all possible subsets. Here the amount of uncertainty due to randomness is much more than m_5; but the *average* amount of uncertainty due to nonspecificity is nearly the same as that of m_5. Hence the total uncertainty for m_6 is expected to be greater than for m_5. Apparently m_7 represents the case of maximum uncertainty. In m_6 the basic assignment function is uniformly distributed on $P(X)$. m_7 also concentrates on all elements of $P(X)$ but the belief value attached to a set is proportional to its cardinality. This increases the nonspecificity of m_7, which distributes both randomness and nonspecificity uniformly over the largest possible set of focal elements. What happens in Example 3 when TU is computed using equations (12), (15) and (21)?

The second and third columns of Table 2 list the values of global uncertainty as suggested by Lamata and Moral, whereas the fourth column displays the values of total uncertainty as given by the function of Klir and Ramer for the seven BPAs of Example 3. Table 2 reveals that none of these measures conforms to our intuitive desire regarding the inequalities $TU(m_1) < TU(m_2) = TU(m_3) < TU(m_4) < TU(m_5) < TU(m_6) < TU(m_7)$ about *total* uncertainty. Note also that G_2 and T are identical on all seven BPAs, so in this example there is no quantitative difference

Table 2. Global and Total Uncertainty Values for the BPAs in Example 3

m	G_1 via (12)	G_2 via (15)	T via (21)
m_1	0.0	0.0	0.0
m_2	2.0	2.0	2.0
m_3	2.0	2.0	2.0
m_4	1.415	2.0	2.0
m_5	1.263	2.0	2.0
m_6	1.353	2.0	2.0
m_7	1.394	2.0	2.0

between G_2 and T, even though there is an ostensible qualitative difference in what they measure.

6. COMPUTATIONAL COMPLEXITY OF T AND G_1

Computational problems in the framework of the theory of evidence are well known [17, 18]. This section compares the computational complexity of T and G_1 when only the basic assignment function, m is available. We assume that $|F| = N$, ie, m focuses on N subsets. The total number of logarithmic evaluations, multiplications and additions (the exact number might vary depending on implementation) are summarized in Table 3 (totals are computed ignoring extra overhead in Logarithm evaluation).

We have not considered G_2 because the total number of operations for it is a function of n, the cardinality of $|X|$, whereas for T and G_1 it is a function of N. Hence it is difficult to compare the computational overhead of G_2 to that of T and G_1. Moreover, it is difficult to find an expression for the total computations involved in an evaluation of G_2. From Table 3 one sees that the computational overhead for both G_1 and T is substantial. The number of logarithmic evaluations are the same for both functions. If we assume that multiplication and addition require the same amount of time, then G_1 and T involve $(N^2 + 2N - 1)$ and $(3N^2 + 2N - 1)$ additions, respectively. Hence, G_1 and T are both $O(N^2)$ procedures. Thus, when the number of possibilities is n, N may be as high as $2^n - 1$, indicating excessive computation for both G_1 and T.

7. CONCLUSIONS AND DISCUSSION

Limitations of some existing measures of the probabilistic and nonspecific components of non-fuzzy uncertainty that have been used in evidential reasoning were examined. The measures of conflict discussed in section 3 cannot account for complete uncertainty (ignorance) that arises

Table 3. Time Complexity of T and G_1 ($|F| = N$, $|X| = n$)

Computation	T via (21)	G_1 via (12)
Logarithmic	N	N
Multiplication	$2N^2 + 2N$	$2N$
Addition	$N^2 - 1$	$N^2 - 1$
Total	$3N^2 + 3N - 1$	$N^2 + 3N - 1$

due to randomness. We have stated and proved several new theorems on various nonspecificity measures.

We discussed three composite measures of total and global uncertainty. Lamata and Moral defined a global measures of uncertainty (G_1) as the sum of Yager's measure of dissonance (E) and the nonspecificity measure (I) of Dubois and Prade. They also defined a second measure of global uncertainty (G_2) as the sum of a measure of innate contradiction (V) and a measure of imprecision (W). Klir and Ramer defined total uncertainty (T) as the sum of I and a new measure of conflict (D). We have established by examples and theorems that all of these composite measures result in intuitively unappealing situations. None of G_1, G_2 nor T has a unique maximum; hence, it is difficult to interpret (visualize) the ambiguity present in the system reflected by values of G_1, G_2 and T. G_1 and T do quantify some interesting aspects of uncertainty, but the name *total* (or *global*) uncertainty does not seem appropriate for either of them, because conflict and nonspecificity represent different aspects of non-fuzzy uncertainty, so there is no sound rationale for simply adding them together to assess the total uncertainty of a *BPA*. Computational overhead, an important consideration in evidential reasoning, may be quite large for G_1 and T.

Since the probabilistic and nonspecific components of total uncertainty are coupled in an unknown manner it seems better to look at total uncertainty as a *whole*, rather than as a sum of conflict and nonspecificity. Along with all of the problems of composite measures itemized above, this fact motivates us to look for a new measure of average total uncertainty. Our approach in Part II of this paper will be to postulate a set of axioms that seem desirable for any measure of total uncertainty (as opposed to axioms related to only one component in a composite sum); and then to derive a function that satisfies these axioms.

ACKNOWLEDGMENT

The authors gratefully acknowledge discussions with Professors G. Klir, A. Ramer, H. Nguyen, P. Smets, H. Prade, and D. Dubois, each of whom has helped us to improve these papers significantly.

References

1. Klir, G. J., and Folger, T. A., *Fuzzy Sets, Uncertainty, and Information*, Prentice Hall, N.J., 1988.
2. Pal, N. R., and Pal, S. K., Higher order fuzzy entropy and hybrid entropy of a set, *Inf. Sci.* 61(3), 211–231, 1992.

3. Deluca, A., and Termini, S., A definition of nonprobabilistic entropy in the setting of fuzzy sets theory, *Inf. Control.* 20, 301–312, 1972.

4. Xie, W. X., and Bedrosian, S. D., An information measure for fuzzy sets, *IEEE Trans. Syst. Man. Cybern.* SMC-14(1), 151–156, 1984.

5. Shafer, G., *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, N.J., 1976.

6. Yager, R. R., Entropy and specificity in a mathematical theory of evidence, *Int. J. Gen. Syst.* 9, 249–260, 1983.

7. Hohle, U., Fuzzy plausibility measures, in E. P. Klement (ed.), *Proceedings of the 3rd International Seminar on Fuzzy Set Theory*, Johannes Kepler University, Linz, 7–30, 1981.

8. Hohle, U., Entropy with respect to plausibility measures, *Proceedings of the 12th IEEE International Symposium on Multiple Valued Logic*, Paris, 167–169, 1982.

9. Smets, P., Information content of an evidence, *Int. J. Man-Mach. Stud.* 19, 33–43, 1983.

10. Higashi, M., and Klir, G. J., Measures of uncertainty and information based on possibility distributions, *Int. J. Gen. Syst.* 9, 43–58, 1983.

11. Dubois, D., and Prade, H., A note on measures of specificity for fuzzy sets, *Int. J. Gen. Syst.* 10, 279–283, 1985.

12. Klir, G. J., and Ramer, A., Uncertainty in Dempster–Shafer theory: A critical re-examination, *Int. J. Gen. Syst.* 18, 155–166, 1990.

13. Lamata, M. T., and Moral, S., Measures of entropy in the theory of evidence, *Int. J. Gen. Syst.* 14, 297–305, 1987.

14. Dubois, D., and Prade, H., A set theoretic view of belief functions: logical operations and approximations by fuzzy sets, *Int. J. Gen. Syst.* 12, 193–226, 1986.

15. Yager, R. R., Measuring tranquility and anxiety in decision making: An application of fuzzy sets, *Int. J. Gen. Syst.* 8, 139–146, 1982.

16. Dubois, D., and Prade, H., Properties of measures of information in evidence and possibility theories, *Fuzzy Sets Syst.*, 24, 161–182, 1987.

17. Shafer, G., and Pearl, J., Readings in Uncertain Reasoning, Morgan Kauffmann, Palo Alto, C.A., 1990.

18. Barnett, J. A., Computational methods for a mathematical theory of evidence, *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, Vancouver, Canada, 868–875, 1981.

19. Klir, G. J., and Parviz, B., A note on the measure of discord, Submitted to 8th Conference Uncertainty in AI, Stanford, 1992 (personal communication).