Uncertainty Measures for Evidential Reasoning II: A New Measure of Total Uncertainty

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

In Part I we discussed limitations of two measures of global (non-fuzzy) uncertainty of Lamata and Moral, and a measure of total (non-fuzzy) uncertainty due to Klir and Ramer and established the need for a new measure. In this paper we propose a set of intuitively desirable axioms for a measure of total uncertainty (TU) associated with a basic assignment m(A), and then derive an expression for a (unique) function that satisfies these requirements. Several theorems are proved about the new measure. Our measure is additive, and unlike other TU measures, has a unique maximum. The 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. We show that the computational complexity of the new measure is $O(N)$, *whereas previous measures of TU are* $O(N^2)$. *Finally, we compare the new measure to its predecessors by extending the numerical example of Part I so that it includes values of the new measure.*

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

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.

This research was supported by NSF IRI-9003252.
Received December 1, 1991; accepted May 1, 1992.

International Journal of Approximate Reasoning 1993; 8:1–16

655 Avenue of the Americas, New York, NY 10010 0888-613X/93/$5.00

1. INTRODUCTION

In Part I we noted that uncertainty can be broadly divided into two facets: fuzzy and non-fuzzy. Fuzzy uncertainty arises when the boundary of a set is not crisply defined, or when an event can partially occur; for example, when the output of a die-throwing experiment is described by linguistic hedges such as *high* or *low*. The literature is quite rich on measures of fuzzy uncertainty; interested readers may refer to [1–4]. On the other hand, non-fuzzy uncertainty arises when there is randomness and/or nonspecificity associated with a system. Referring to the same example of a die, the outcome could be described as one of the six faces. Here the system output is unambiguously (crisply) defined. The present study confines itself only to measures of non-fuzzy uncertainty.

We observed in Part I that many writers consider the total *nonfuzzy* uncertainty to comprise at least two components: uncertainty due to nonspecificity and to randomness. Several authors have suggested different measures for these two aspects of uncertainty, [6–17] many of which were reviewed in Part I. Yager [6] proposed a measure called dissonance or conflict, whereas Hohle [7, 8] suggested a measure to quantify the level of confusion present in a body of evidence. Higashi and Klir [9] proposed a measure of nonspecificity for a possibility distribution that was later extended to any body of evidence by Dubois and Prade [10]. Recently Klir and Ramer [11] pointed out some limitations of the measure of conflict (confusion) of Hohle [7, 8] and suggested a new measure for the same. Lamata and Moral [14] proposed two composite measures, called *global* uncertainty measures. Klir and Ramer [11] suggested another composite measure called *total* uncertainty that also defined *TU* as the sum of nonspecificity and a new measure of conflict they called discord. In Part I we discussed several problems that may arise when different elementary measures are added to form a composite measure for total uncertainty. Because elementary measures such as dissonance, discord, or nonspecificity are derived using different objectives—each represents a different facet of uncertainty—the interpretation of composite measures such as the *total* or *global* uncertainties alluded to above becomes difficult. Aggregation of different elementary measures depends on the mode of interaction among the various aspects of non-fuzzy uncertainty represented by elementary factors. Because nonspecificity and randomness are related in an unknown way, we argued in Part I that a more rational approach to the quantification of total non-fuzzy uncertainty would be to hypothesize a set of axioms that captured intuitively desirable properties of *TU*; and then try to derive a function that satisfies the axioms, rather than add together possibly conflicting elementary components. The measure we propose in this paper is thus aimed at assessing the total uncertainty arising in a body of evidence due to both randomness (ignorance and inconsistency) and nonspecificity associated with a basic assignment function.

2. A NEW MEASURE OF TOTAL UNCERTAINTY

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 equations (1) and (2) of Part I. 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 (ie, 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.

Let M_n denote the set of all BPAs on the power set with 2^n elements. A measure of uncertainty is a mapping S: $M_n \to [0, \infty)$ that captures 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 simple multiplication by a constant, we shall omit the base unless clarity demands it.

Our discussion of different uncertainty measures in Part I indicates two things: first, the uncertainty associated with $m(A)$ should be inversely related to the value of $m(A)$; second, it should increase with the cardinality of focal elements. In other words, if $m(A) = p$, $m(B) = q$, $|A| = |B|$, and $p > q$, then the total ambiguity associated with $m(B)$ is more than that associated with $m(A)$. And on the other hand, if $m(A) = m(B) = k$ and $|A| > |B|$, the uncertainty should be greater for $m(A)$ because A contains more elements than B. These intuitive requirements lead us to propose some desirable properties for a new measure of total uncertainty.

Let φ: $(0, 1] \to [0, \infty)$ be any function that measures the uncertainty associated with a basic assignment $m(A) > 0$. If φ is to measure the total uncertainty associated with $m(A)$, we feel that it should satisfy each of the following requirements:

- *TU1*: $\varphi(m(A)) \geq 0$.
- *TU2*: $\varphi(m(A))$ should increase with $|A|$.
- *TU3*: $\varphi(m(A))$ should decrease if $m(A)$ increases.
- *TU4*: (Additivity). Suppose m_1 and m_2 are *BPA*s on X and Y, respectively and m is the joint *BPA* on $X \times Y$. If m_1 and m_2 are strongly independent (cf. below) then $\varphi(m(A \times B)) = \varphi(m_1(A)) + \varphi(m_2(B))$ for all $A \subseteq X$, $B \subseteq Y$ and $A \times B \subseteq X \times Y$.
- *TU5*: (Continuity). $\varphi(m(A))$ should be a continuous function of $m(A)$.
- *TU6*: (Normalization). If $m(A) = 1$ and $|A| = 2$, then $\varphi(m(A)) = 1$.
- *TU7*: (Minimum). If $m(A) = 1$ and $|A| = 1$, then $\varphi(m(A)) = 0$.

Note especially that *TU3* does *not* mean that, for a given A, the average or total measure of uncertainty decreases with an increase of $m(A)$. Indeed, an increase in $m(A)$ implies a corresponding decrease of some other value $m(B)$ that is accounted for by the averaging process.

We say m_1 and m_2 are *strongly independent* if and only if $m(A \times B) = m_1(A)m_2(B)$. Thus *TU4* requires that $\varphi(xy) = \varphi(x) + \varphi(y)$. It is well known that the only continuous function (not identically zero) satisfying $\varphi(m_1(A)m_2(B)) = \varphi(m_1(A)) + \varphi(m_2(B))$ is the logarithmic function [18]. Increasing functions of $|A|/m(A)$ satisfy *TU2* and *TU3*. Further, *TU2* and *TU3*, together with *TU4*, force φ to have the form $\varphi(m(A)) = Log_b(|A|/m(A))$ for $b > 1$. Thus $Log_b(|A|/m(A))$ (where b is the base of logarithm) is a function that satisfies properties *TU2–TU5*. Because $|A|/m(A) \geq 1$, this function also satisfies property *TU1*. In order to satisfy *TU6*, one has to take $b = 2$. Hence, the only function that satisfies properties *TU1–TU7* is the function $\varphi(m(A)) = Log_2\{|A|/m(A)\}$. Thus motivated, we define the *expected value* of $\varphi(m(A)) = Log_2\{|A|/m(A)\}$ as the *average total uncertainty* of *BPA* m as:

$$H(m) = \sum_{A \in F} m(A) Log_2\{|A|/m(A)\} \tag{1}$$

or
$$H(m) = \sum_{A \in F} m(A) Log_2|A| - \sum_{A \in F} m(A) Log_2 m(A) \tag{2}$$

or, recalling the expression for I from Table 1 in Part I;

$$H(m) = -\sum_{A \in F} m(A) Log_2 m(A) + I(m) \tag{3}$$

$$= G(m) + I(m), \tag{4}$$

where $G(m) = -\sum_{A \in F} m(A) Log_2 m(A)$.

Comparing H to the composite measures G_1 and T [Table 1, below] shows that our definition of average total uncertainty can also be "decomposed" in the sense that it incorporates Dubois and Prade's nonspecificity measure I and another term. The term $G(m)$ is in some sense analogous to Yager's $(E(m))$ or Klir and Ramer's $(D(m))$ measures of conflict. If viewed independently as an elementary measure, $G(m)$ appears to gauge the inconsistency and mistrust (lack of confidence) associated with a body of evidence. As the randomness of the body of evidence increases, so does the probabilistic uncertainty, while our confidence decreases. Noting the similarity to Shannon's entropy function, $G(m)$ could also be interpreted (again if viewed independently) as a measure of scattering of the evidence among various focal elements. When the body of evidence is more scattered, ignorance is greater. Note that these two interpretations of G do not contradict one another. In this regard we mention that Nguyen [19] has investigated the function G when m is a probability measure in the context of the entropy of random sets. In the present case m is not necessarily a probability measure.

For an *ordered* possibility distribution H takes the form:

$$H(m) = \sum_{i=1}^{n} (\rho_i - \rho_{i+1}) Log_2 i - \sum_{i=1}^{n} (\rho_i - \rho_{i+1}) Log_2 (\rho_i - \rho_{i+1}) \quad (5)$$

In the next section we develop a theoretical foundation that justifies the use of H as a measure of TU. We close this section by exhibiting a new version of Table 1, Part I that has been revised to include the measure H and its "components." We emphasize again that although the components G and I of H shown in the upper part of this table suggest that H is a composite measure analogous to G_1, G_2 and T, in fact H arises directly from a desire to find a measure satisfying *TU1–TU7* and it just turns out that H has the two components shown in Table 1.

3. SOME PROPERTIES OF H

Our first result shows that H is actually stronger than the requirements of *TU7*.

THEOREM 1 *$H(m) = 0 \Leftrightarrow m(A) = 1$ for some A such that $|A| = 1$.*

Proof If $m(A) = 1$ and $|A| = 1$, then for all B such that $B \neq A$, $m(B) = 0$. This is by definition of m; hence $H(m) = 0$.

Now suppose $H(m) = 0$. Then $|A|/m(A) = 1$ for all $A \in F$, as $m(A) Log_2\{|A|/m(A)\} \geq 0$. Now because $m(A) \in (0, 1]$ and $|A| \geq 1$, $|A|/m(A) = 1$ only when $m(A) = 1$ and $|A| = 1$. By definition of m, if $m(A) = 1$, then $m(B) = 0$ for all $B \neq A$. ∎

Table 1. A Catalog of Uncertainty Measures—Revised

Author(s)	Sum	Probabilistic	Nonspecific
Yager [6]		$E(m) = -\sum_{A \in F} m(A) Log Pl(A)$ (Dissonance)	$J(m) = 1 - \sum_{A \in F} \{m(A)/\|A\|\}$ (Nonspecificity)
Hohle [7, 8]		$C(m) = -\sum_{A \in F} m(A) Log Bel(A)$ (Confusion)	
Higashi & Klir [9]			$U(r) = \sum_{i=1}^{n} m(A_i) Log\|A_i\|$ (Nonspecificity)
Smets [16]		$L(m) = -\sum_{A \subseteq X} Log Cm(A)$	
Dubois & Prade [10]			$I(m) = \sum_{A \in F} m(A) Log\|A\|$ (Nonspecificity)
Klir & Ramer [11]		$D(m) = -\sum_{A \in F} m(A) Log\left[\sum_{B \in F} m(B)\|A \cap B\|/\|B\|\right]$ (Discord)	
Lamata & Moral [14]		$V(m) = EV_{Bel}(-Log(Pl(\{x\})))$ (Innate Contradiction)	$W(m) = Log\left(\sum_{A \subseteq X} m(A)\|A\|\right)$ (Imprecision)
Pal, Bezdek, & Hemasinha		$G(m) = -\sum_{A \in F} m(A) Log_2 m(A)$	$I(m) = \sum_{A \in F} m(A) Log\|A\|$ (Nonspecificity)
Lamata & Moral [14]	$G_1(m)$	$= E(m) + I(m)$ (global uncertainty)	
Lamata & Moral [14]	$G_2(m)$	$= V(m) + W(m)$ (global uncertainty)	
Klir & Ramer [11]	$T(m)$	$= D(m) + I(m)$ (total uncertainty)	
Pal, Bezdek, & Hemasinha	$H(m)$	$= G(m) + I(m)$ (average total uncertainty)	

Next, we show that H becomes Shannon's entropy when the belief structure is a probabilistic model.

THEOREM 2 *$H(m)$ reduces to Shannon's entropy when m represents a probability distribution.*

Proof If m is a probability distribution then m focuses only on singleton elements. In other words, $m(A) = 0$ if $|A| > 1$ and $\Sigma m(\{x_i\}) = 1$. Thus for a probability distribution $I(m) = 0$ and

$$H(m) = G(m) = -\sum m(\{x_i\}) Log_2 m(\{x_i\}) = -\sum p(x_i) Log_2 p(x_i).$$

■

At the other extreme we have the following relationship between H and Hartley's measure of information.

THEOREM 3 *$H(m)$ reduces to Hartley's Entropy if $m(X) = 1$.*

Proof If $m(X) = 1$ then by definition of m, $m(A) = 0$ *if* $A \neq X$. Thus $I(m) = log_2|X|$ and $G(m) = 0$. Hence $H(m) = log_2|X|$, which is Hartley's entropy. ■

Now suppose m_1 and m_2 are *BPA*s on X and Y, respectively, and m is the joint basic probability assignment on $X \times Y$. Recall that m_1 and m_2 are *strongly independent* if $m(A \times B) = m_1(A)m_2(B)$ for all $A \subseteq X$, $B \subseteq Y$.

THEOREM 4 *H is additive; ie, for any strongly independent m on $X \times Y$, $H(m) = H(m_1) + H(m_2)$.*

Proof Let $R = A \times B \subseteq X \times Y$. We calculate:

$$\begin{aligned} H(m) &= \sum_{R \subseteq X \times Y} m(R) Log_2|R| - \sum_{R \subseteq X \times Y} m(R) Log_2 m(R) \\ &= \sum_{A \subseteq X,\, B \subseteq Y} m(A \times B) Log_2|A \times B| \\ &\quad - \sum_{A \subseteq X,\, B \subseteq Y} m(A \times B) Log_2 m(A \times B) \\ &= \sum_{A \subseteq X,\, B \subseteq Y} m_1(A) m_2(B) Log_2|A|\,|B| \\ &\quad - \sum_{A \subseteq X,\, B \subseteq Y} m_1(A) m_2(B) Log_2(m_1(A) m_2(B)) \\ &= \sum_{A \subseteq X,\, B \subseteq Y} m_1(A) m_2(B) Log_2|A| \end{aligned}$$

$$+ \sum_{A \subseteq X,\, B \subseteq Y} m_1(A) m_2(B) Log_2|B|$$

$$- \sum_{A \subseteq X,\, B \subseteq Y} m_1(A) m_2(B) Log_2(m_1(A))$$

$$- \sum_{A \subseteq X,\, B \subseteq Y} m_1(A) m_2(B) Log_2(m_2(B))$$

$$= \left(\sum_{B \subseteq Y} m_2(B) \right) \sum_{A \subseteq X} m_1(A) Log_2|A|$$

$$+ \left(\sum_{A \subseteq X} m_1(A) \right) \sum_{B \subseteq Y} m_2(B) Log_2|B|$$

$$- \left(\sum_{B \subseteq Y} m_2(B) \right) \sum_{A \subseteq X} m_1(A) Log_2 m_1(A)$$

$$- \left(\sum_{A \subseteq X} m_1(A) \right) \sum_{B \subseteq Y} m_2(B) Log_2 m_2(B)$$

$$= \sum_{A \subseteq X} m_1(A) Log_2|A| + \sum_{B \subseteq Y} m_2(B) Log_2|B|$$

$$- \sum_{A \subseteq X} m_1(A) Log_2 m_1(A) - \sum_{B \subseteq Y} m_2(B) Log_2 m_2(B)$$

$$= \sum_{A \subseteq X} m_1(A) Log_2|A| - \sum_{A \subseteq X} m_1(A) Log_2 m_1(A)$$

$$+ \sum_{B \subseteq Y} m_2(B) Log_2|B| - \sum_{B \subseteq Y} m_2(B) Log_2 m_2(B)$$

$$= H(m_1) + H(m_2). \qquad \blacksquare$$

When the basic assignment function is most uniformly (ambiguously) distributed over all possible subsets of X, total uncertainty should attain a maximum. In other words, when $m(A) > 0$ and $m(A)/|A| = m(B)/|B|$ for all non-empty subsets A and B of X, total uncertainty should maximize. The quantity $m(A)/|A|$ can be viewed as the degree of belief attached per possible choice in A. Theorem 5 exhibits the unique global maximum of H.

THEOREM 5 *$H(m)$ attains its unique global maximum when $m(A) = |A|/k$, where $k = \sum_{i=1}^{n} i \binom{n}{i} = n2^{n-1}$.*

Proof (For convenience and without loss, we use the natural logarithm in formulation of the proof):

Let $k_i = \binom{n}{i}$, $i = 1, 2, \ldots, n$. If $A_{i,1}, \ldots, A_{i,k_i}$ are the k_i distinct subsets of cardinality i then for $j = 1, 2, \ldots, k_i$ denote $m(A_{i,j})$ by $s_{i,j}$. Thus the problem of maximizing H over M_n reduces to

$$\text{maximize} \quad \left\{ H(S) = H(s_{1,1}, s_{1,2}, \ldots, s_{n,1}) = \sum_{i=1}^{n} \sum_{j=1}^{k_i} s_{i,j} Log(i/s_{i,j}) \right\} \quad \text{subject to}$$

$$0 \le s_{i,j} \le 1, i = 1, \ldots, n; \; j = 1, 2, \ldots, k_i \tag{6a}$$

$$\text{and} \quad \sum_{i=1}^{n} \sum_{j=1}^{k_i} s_{i,j} = 1 \tag{6b}$$

First we show that $H(m)$ (or equivalently $H(S)$) is a strictly concave function over the convex region defined by constraints (6). Let $H(s) = \Sigma h(s_{i,j})$ where the summation is taken over all i and j and $h(s_{i,j}) = s_{i,j} Log\, i/s_{i,j}$. The function $h(t) = t\,Log(i/t)$, $1 \le i \le n$ and $0 \le t \le 1$ is a strictly concave function over $[0, 1]$ because $h''(t) = -1/t < 0$ for $0 < t \le 1$. Because the sum of a finite set of strictly concave functions is again strictly concave, H is strictly concave over the feasible region. Because H is a continuous, strictly concave function, and the feasible region is compact and convex, H has a unique global maximum over the feasible region.

In order to find the optimal point we use the Lagrange multiplier method. At first we ignore constraint (6a), ie, $0 \le s_{i,j} \le 1$, and find an optimal solution. Then we show that (6a) is satisfied by the solution. The Lagrangian formulation follows:

$$\text{Maximize} \quad G(S) = \sum_{i=1}^{n} \sum_{j=1}^{k_i} s_{i,j} Log(i/s_{i,j}) - \lambda \left(\sum_{i=1}^{n} \sum_{j=1}^{k_i} s_{i,j} - 1 \right) \tag{7}$$

where λ is the Lagrangian multiplier needed to account for (6b). Differentiating G with respect to $s_{i,j}$ for any i and j, $1 \le i \le n$ and $1 \le j \le k_i$ and setting G' to zero yields the following necessary conditions for an extremum:

$$Log(i) - 1 - Log(s_{i,j}) - \lambda = 0,$$

$$\text{Thus} \quad s_{i,j} = i/e^{1+\lambda}, \qquad 1 \le i \le n; \quad 1 \le j \le k_i.$$

The other necessary condition for extremality is given by (6b), which can be obtained by differentiating G with respect to λ. Now substituting the value of $s_{i,j}$ in (6b) we get:

$$\sum_{i=1}^{n}\sum_{j=1}^{k_i} s_{i,j} = \sum_{i=1}^{n}\binom{n}{i}(i/e^{1+\lambda}) = 1 \Rightarrow n2^{n-1}/e^{1+\lambda} = 1$$

$$\Rightarrow e^{1+\lambda} = n2^{n-1} \Rightarrow \lambda = Log(n2^{n-1})\text{-}1.$$

Substituting the value of λ in the expression for $s_{i,j}$, one gets:

$$*s_{i,j} = i/(n2^{n-1}), \qquad 1 \le i \le n; \quad 1 \le j \le k_i. \tag{8}$$

Note that $*S = (*s_{i,j})$ given by equation (8) not only satisfies constraint (6b) but also satisfies (6a), that is, $0 \le s_{i,j} \le 1$. Thus $H(S)$ attains an extremum over the feasible region at $*S$. Moreover, since $H(S)$ is a strictly concave function and $\nabla H(*S)(S - *S) = 0$ for any S in the convex feasible region, $*S$ is the unique global maximum point of $H(S)$ over the feasible region [20] ($\nabla H(S)$ is the gradient of H at S). Hence H attains its global maximum at $s_{i,j} = i/(n2^{n-1})$, (ie, at $m(A) = |A|/n2^{n-1}$). ■

Note that in Shannon's information theory, the maximum number of bits required to represent the uncertainty associated with an n-state system is $Log_2 n$, but here it is more than that. In the probabilistic (Shannon's) framework the complete status of a system can be described (ie, total uncertainty can be resolved) by $Log_2 n$ bits, because the status of the system is completely representable only by the state number. This should not be confused with the present case. In the present case the basic assignment can focus on several (up to $2^n - 1$) subsets and each subset can have non-empty intersection with many others. In this case, the total number of possible hypotheses is $2^n - 1$. Moreover, in addition to randomness, the total uncertainty includes nonspecificity. Therefore, one can not even say that the total uncertainty should be bounded (above) by $Log_2(2^n - 1)$. In this context, we mention that there are several [21] probabilistic entropy measures for which the maximum value exceeds $Log_2 n$. For example, Kapur's entropy of the fifth kind has the maximum $n^{(1-\alpha)/\alpha}$, where $0 < \alpha < 1$ [21].

THEOREM 6 *For an ordered possibility distribution r of length n*, $H(r) = 0$ $\Leftrightarrow$ *r is the smallest possibility distribution of length n.*

Proof The smallest ordered possibility distribution of length n is given by

$$r(x_i) = \begin{Bmatrix} 1, & \text{if } i = 1 \\ 0, & \text{otherwise} \end{Bmatrix}.$$

Thus for the smallest possibility distribution of length n, the basic probability assignment function is given by

$$m(A_i) = \begin{cases} 1, & \text{if } i = 1 \\ 0, & \text{otherwise} \end{cases},$$

where $A_i = \{x_1, x_2, \ldots, x_i\}$. Hence for the smallest possibility distribution $H(r) = 0$. On the other hand, for an ordered possibility distribution r of length n, $H(r)$ is defined as:

$$\begin{aligned} H(r) &= \sum_{i=1}^{n} m(A_i) Log_2 |A_i| - \sum_{i=1}^{n} m(A_i) Log_2 m(A_i) \\ &= \sum_{i=1}^{n} m(A_i) Log_2 i - \sum_{i=1}^{n} m(A_i) Log_2 m(A_i) \\ &= \sum_{i=1}^{n} m(A_i) Log_2 \{i/m(A_i)\}. \end{aligned}$$

Now $m(A_i) Log_2\{i/m(A_i)\} \geq 0$ for all $i = 1, \ldots, n$. Therefore,

$$\begin{aligned} H(r) = 0 &\Rightarrow m(A_i) Log_2\{i/m(A_i)\} = 0 \quad \text{for all } i. \\ &\Rightarrow \text{either} \quad Log_2\{i/m(A_i)\} = 0 \quad \text{or} \quad m(A_i) = 0 \quad \text{for all } i. \\ &\Rightarrow \text{either} \quad \{i/m(A_i)\} = 1 \quad \text{or} \quad m(A_i) = 0 \quad \text{for all } i. \\ &\Rightarrow m(A_1) = 1 \quad \text{and} \quad m(A_i) = 0 \quad \text{for } i > 1, \quad \text{as } \Sigma m(A_i) = 1. \\ &\Rightarrow r(x_i) = \begin{cases} 1, & \text{if } i = 1 \\ 0, & \text{otherwise} \end{cases}. \end{aligned}$$

Hence $H(r) = 0$ if and only if r is the smallest possibility distribution. ∎

THEOREM 7 *For the largest ordered possibility distribution,* $H(m) = Log_2 n$.

Proof The largest ordered possibility distribution is defined as $r(x_i) = 1$ for all $i = 1, \ldots, n$. In other words, $m(A_n) = m(X) = 1$ and $m(A_i) = 0$ if $i \neq n$. Direct substitution of the above equations in the expression for H shows $H(m) = Log_2 n$. ∎

Finally, we show the relationship between H and the previous composite measures of global or total uncertainty called G_1 and T in Table 1. The following theorem applies to our discussion made in Section 3 of Part I about the failure of G_1 and T to capture the total amount of ignorance that arises due to randomness.

THEOREM 8 $G_1 \le T \le H$.

Proof Because $G_1 = E + I$, $T = D + I$, and $H = G + I$, it suffices to show that $E \le D \le G$. Klir and Ramer [11] have shown that $E \le D$. It remains to be seen that $D \le G$. Towards this end, consider the following expression:

$$\sum_{B \in F} m(B)|A \cap B|/|B| = m(A)|A \cap A|/|A| + \sum_{\substack{B \in F \\ B \neq A}} m(B)|A \cap B|/|B| \ge m(A)$$

hence, $Log_2(\Sigma_{B \in F} m(B)|A \cap B|/|B|) \ge Log_2 m(A)$

$$\Rightarrow m(A)Log_2\left(\sum_{B \in F} m(B)|A \cap B|/|B|\right) \ge m(A)Log_2 m(A)$$

$$\Rightarrow -\sum_{A \in F} m(A)Log_2\left(\sum_{B \in F} m(B)|A \cap B|/|B|\right) \le -\sum_{A \in F} m(A)Log_2 m(A)$$

$$\Rightarrow D \le G$$ ■

4. A NUMERICAL COMPARISON OF G_1, G_2, T, AND H

We investigate the characteristics of H on the BPAs of Example 3, Part I, the values of which are repeated here for the convenience of readers:

EXAMPLE 3 (cf. Part I) 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 $A \in P(X)$, $A \neq \varnothing$.
- $\mathbf{m}_7$:

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

Column five of Table 2 exhibits values of H for these *BPA*s. This is an enlarged version of Table 2 in Part I that displays values of H as well as those G_1, G_2, and T. From the table we see that $G_1 = G_2 = T = H$ on m_1, m_2, and m_3. It is interesting to note that G_1 and H both increase as one moves from m_5 to m_7, which is indeed desirable. However, the maximum value of G_1 occurs at m_2 and m_3. T and G_2 do not change at all as we move from m_2 to m_7. This behavior of T and G_2 makes it difficult to interpret their usefulness as measures of total uncertainty. On the other hand, H increases as the *BPA* goes from m_3 to m_7 and satisfies all the intuitive properties discussed in Part I. Note that G_1, G_2, and $T \leq H$ for all m_i, $i = 1, \ldots, 7$. This agrees with our explanation in Section 3, Part I, that G_1, G_2, and T do not account for the complete uncertainty due to randomness.

5. COMPUTATIONAL COMPLEXITY OF G_1, G_2, T, AND H

Computational overhead associated with some of the existing composite measures of *TU* was discussed and illustrated in connection with Table 3 of Part I. Table 3 below is an enhanced version of that table which offers a comparison of the computational complexities of H with T and G_1 when only the basic assignment function, m is available. (Totals are computed ignoring the extra overhead involved in Logarithmic evaluations). It is assumed that $|F|$ is equal to N, ie, m focuses on N subsets and $|X| = n$.

Table 3 shows that the computational overhead is minimum for H and maximum for the composite measure T proposed by Klir and Ramer [11]. 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. On the other hand, H requires only $3N - 1$ additions. For example, if m focuses on 100 subsets of X, then computation of T,

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

m	G_1 via (I-1)	G_2 via (I-15)	T via (I-21)	H via (II-2)
m_1	0.0	0.0	0.0	0.0
m_2	2.0	2.0	2.0	2.0
m_3	2.0	2.0	2.0	2.0
m_4	1.415	2.0	2.0	3.0
m_5	1.263	2.0	2.0	3.585
m_6	1.353	2.0	2.0	4.863
m_7	1.394	2.0	2.0	5.0

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

Computation	T via (I-21)	G_1 via (I-12)	H via (II-2)
Logarithmic	N	N	N
Multiplication	$2N^2 + 2N$	$2N$	$2N$
Addition	$N^2 - 1$	$N^2 - 1$	$N - 1$
Total	$3N^2 + 3N - 1$	$N^2 + 3N - 1$	$4N - 1$

G_1, and H require 30199, 10199, and 299 additions, respectively. Computation of E and D (and thus T and G_1) requires maintenance of complex data structures that have been ignored in this analysis. Finally, we observe that G_1 and T are both $O(N^2)$ procedures, whereas, H is $O(N)$.

6. CONCLUSIONS AND DISCUSSION

Limitations of some composite measures of total non-fuzzy uncertainty (with factors accounting for both probabilistic and nonspecific components of uncertainty) used in evidential reasoning have been examined in Part I of this paper [22]. Lamata and Moral defined one composite measure of global uncertainty (G_1) as the sum of Yager's measure of dissonance (E) and the nonspecificity measure (I) of Dubois and Prade. Klir and Ramer defined another measure of total uncertainty (T) as the sum of I and a new measure of conflict (D). We established in Part I that these composite measures result in intuitively unappealing situations and suggested that this situation called for a new approach to the measurement of total uncertainty due to a *BPA*. To achieve this a set of (intuitively) desirable axioms for such a measure of average total uncertainty was proposed in Part II. Based on these axioms we have derived a new measure (H) and have proven several theorems about it.

The new measure, albeit based on a noncomposite approach to *TU*, leads to a function that can be factored into the sum of Dubois and Prade's nonspecificity and an entropy-like measure associated with random sets that was previously studied by Nguyen. Under complete ignorance, the new measure reduces to Hartley's information. On the other hand, when the *BPA* concentrates only one singletons, the new measure is equivalent to Shannon's probabilistic entropy. Our measure H has a unique maximum, in sharp contradistinction to the composite measures G_1, G_2, and T. We have shown that H attains its global maximum when the *BPA* distributes both randomness and nonspecificity uniformly over the largest possible set of focal elements. We also proved that $G_1 \leq T \leq H$ over all *BPA*s. Finally, we briefly studied the issue of computational complexity, and showed that H is computationally more tractable than G_1 and T. In fact, both G_1 and T are $O(N^2)$, whereas H is only $O(N)$.

We hope to extend both the theory and practical utility of this new measure of average total uncertainty in a future investigation. For example, one might investigate the possibility of extending this work to the generalized Dempster–Shafer framework [23], and also to the case with an infinite universe of discourse.

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. Higashi, M., and Klir, G. J., Measures of uncertainty and information based on possibility distributions, *Int. J. Gen. Syst.* 9, 43–58, 1983.
10. Dubois, D., and Prade, H., A note on measures of specificity for fuzzy sets, *Int. J. Gen. Syst.* 10, 279–283, 1985.
11. Klir, G. J., and Ramer, A. Uncertainty in Dempster Shafer theory: A critical re-examination, *Int. J. Gen. Syst.* 18, 155–166, 1990.

12. 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.

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

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

15. Dubois, D., and Prade, H., Properties of measures of information in evidence and possibility theories, *Fuzzy sets and systems*, 24, 161–182, 1987.

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

17. Klir, G. J., and Parviz, B. A note on the measure of discord, submitted to *8th Conference Uncertainty in AI*, Stanford (Private communication), 1992.

18. Bartle, R. G., *The Elements of Real Analysis*, John Wiley and Sons, New York, N.Y., 1976.

19. Nguyen, H. T., On entropy of random sets and possibility distributions, in *The Analysis of Fuzzy Information*, Vol. 1, (CRC Press, Boca Raton, FL, 145–156, 1986).

20. Bazaraa, M. S., and Shetty, C. M., *Nonlinear Programming: Theory and Algorithms*, Wiley, New York, N.Y., 1979.

21. Kumar, U., Kumar, V., and Kapur, J. N., Normalized measures of entropy, *Int. J. Gen. Syst.* 12, 55–69, 1986.

22. Pal, N., Bezdek, J., and Hemasinha, R., Uncertainty measures for evidential reasoning I: a review, *Int. J. Approx. Reasoning*, 7(3/4), 165–183, 1992.

23. Yen, J., Generalizing the Dempster–Shafer theory to fuzzy sets, *IEEE Trans. Syst. Man Cybern.* 20(3), 559–570, 1990.