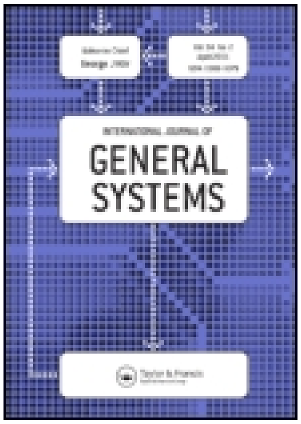

International Journal of General Systems
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ggen20

MEASURES OF UNCERTAINTY AND INFORMATION BASED ON POSSIBILITY DISTRIBUTIONS

MASAHIKO HIGASHI [a] & GEORGE J. KLIR [a]

[a] Department of Systems Science, School of Advanced Technology, State University of New York, Binghamton, New York, U.S.A.

Published online: 31 May 2007.

Int. J. General Systems, 1982, Vol. 9, pp. 43–58
0308-1079/82/0901-0043 $06.50/0

MEASURES OF UNCERTAINTY AND INFORMATION BASED ON POSSIBILITY DISTRIBUTIONS†

MASAHIKO HIGASHI‡ AND GEORGE J. KLIR

Department of Systems Science, School of Advanced Technology, State University of New York, Binghamton, New York, U.S.A.

(Received May 12, 1982; in final form August 12, 1982)

A measure of uncertainty and information for possibility theory is introduced in this paper. The measure is called the U-uncertainty or, alternatively, the U-information. Due to its properties, the U-uncertainty/information can be viewed as a possibilistic counterpart of the Shannon entropy and, at the same time, a generalization of the Hartley uncertainty/information.

A conditional U-uncertainty is also derived in this paper. It depends on the U-uncertainties of the joint and marginal possibility distributions in exactly the same way as the conditional Shannon entropy depends on the entropies of the joint and marginal probability distributions. The conditional U-uncertainty is derived without the use of the notion of conditional possibilities, thus avoiding a current controversy in possibility theory.

The proposed measures of U-uncertainty and conditional U-uncertainty provide a foundation for developing an alternative theory of information, one based on possibility theory rather than probability theory.

INDEX TERMS: Information, uncertainty, probability theory, possibility theory, Shannon entropy, Hartley information, conditional uncertainty, possibility distribution, possibility measure, U-uncertainty.

1 INTRODUCTION

It is well known that the *average amount of information* expected from an experiment which involves a *finite* set of alternative outcomes (events, states, etc.), say set X, can be measured by the *average amount of uncertainty* associated with the set X.[8,9,23,42] However, when a measure of uncertainty is adopted as an information measure, only the syntactic aspects of information, not its semantic and pragmatic aspects, are in fact measured.[14] While such a measure would not be adequate to deal with information in human communication, it is not only adequate but even desirable for dealing with structural (syntactic) aspects of systems.[3–6, 10, 12, 13, 17–20, 28, 39, 41, 42]

Although the measure of information in terms of uncertainty was originally investigated for the purpose of analyzing and designing telecommunication systems,[37] we are interested in its use in a much larger context, as a powerful tool in general system problem solving methodology.[11,29]

Measures of uncertainty have been predominantly studied in terms of probability distributions associated with a relevant set X of n alternative outcomes. In order to describe some of the most important measures in this category, let

$$X=\{x_1, x_2, \ldots, x_n\},$$

let p_i denote the probability of $x_i \in X$ ($i \in \{1, 2, \ldots, n\} = N_n$),† let

$$^nP=\left\{(p_1, p_2, \ldots, p_n) \,\middle|\, \sum_{i=1}^{n} p_i = 1, p_i \geqq 0 \quad \text{for all} \quad i \in N_n\right\}$$

†This work was supported by the National Science Foundation under Grant No. ECS-8006590.

‡Also Department of Biophysics, University of Kyoto, Japan.

†N subscripted by a positive integer is always used in this paper to denote the set of positive integers from 1 through the value of the subscript.

denote the set of all probability distributions associated with n outcomes and let

$$P=\bigcup_{n\in N} {}^{n}P.$$

Then, every measure of uncertainty based on probability distributions is characterized by a function

$$H_M: P\to[0,\infty)$$

which possesses some properties considered desirable for such a measure. The name "*entropy*" is usually used for these measures, a name which was adopted from thermodynamics.[8,9] Symbol H is used here to indicate that the uncertainty measure is based on probability distributions (a symbol usually used in the literature); subscript M distinguishes measures with different properties.

The classical measure of uncertainty H_S, which has dominated the literature on information theory since it was proposed by Shannon in 1948[37] is defined as follows:

$$H_S(p_1,p_2,\ldots,p_n)=-\sum_{i=1}^{n} p_i\log_2 p_i$$

It is called the *Shannon entropy*.

Several classes of entropies have been described in the literature, each of which contains the Shannon entropy as a special case. They include:

1) *Rényi's entropies* H_α (also called entropies of degree α), defined for all real numbers $\alpha\neq 1$ as follows:[1,35]

$$H_\alpha(p_1,p_2,\ldots,p_n)=\frac{1}{1-\alpha}\log_2\sum_{i=1}^{n} p_i^\alpha$$

It is well known that

$$\lim_{\alpha\to 1} H_\alpha(p_1,p_2,\ldots,p_n)=H_S(p_1,p_2,\ldots,p_n).$$

2) *Entropies of order* β, H_β, introduced by Daróczy,[1,21] which have the following form for all $\beta\neq 1$:

$$H_\beta(p_1,p_2,\ldots,p_n)=\frac{1}{2^{1-\beta}-1}\left(\sum_{i=1}^{n} p_i^\beta-1\right)$$

As is the case for Rényi's entropies,

$$\lim_{\beta\to 1} H_\beta(p_1,p_2,\ldots,p_n)=H_S(p_1,p_2,\ldots,p_n)$$

3) *R-norm entropies* H_R, defined for all $R\neq 1$ by the following formula:[7]

$$H_R(p_1,p_2,\ldots,p_n)=\frac{R}{R-1}\left(1-\left(\sum_{i=1}^{n} p_i^R\right)^{1/R}\right)$$

As in the previous classes,

$$\lim_{R\to 1} H_R(p_1,p_2,\ldots,p_n)=H_S(p_1,p_2,\ldots,p_n)$$

Formulas for converting entropies from one class to the other classes are well known. Conversion formulas between the Rényi and Daróczy entropies are derived in Ref. 1, p. 185; formulas for converting the Rényi and Daróczy entropies into the R-norm entropies are given in Ref. 7, p. 144.

There are six properties which are usually considered desirable for a measure of uncertainty, say measure H_M, which is defined in terms of probability distributions:

($H1$) *Symmetry*:[1,2,24]

$$H_M(p_1,p_2,\ldots,p_n)=H_M(p_{b(1)},p_{b(2)},\ldots,p_{b(n)})$$

for all $(p_1,p_2,\ldots,p_n)\in P$ and for all permutations $(b(1),b(2),\ldots,b(n))$ of $(1,2,\ldots,n)$

($H2$) *Expansibility*:

$$H_M(p_1,p_2,\ldots,p_n,0)=H_M(p_1,p_2,\ldots,p_n)$$

for all $(p_1,p_2,\ldots,p_n)\in P$.

($H3$) *Subadditivity*:

$$\begin{aligned}H_M(p_{11},p_{12},\ldots,p_{1n},p_{21},p_{22},\ldots,\\ p_{2n},\ldots,p_{m1},p_{m2},\ldots,p_{mn})\\ \leq H_M\left(\sum_{j=1}^{n} p_{1j},\sum_{j=1}^{n} p_{2j},\ldots,\sum_{j=1}^{n} p_{mj}\right)\\ +H_M\left(\sum_{i=1}^{m} p_{i1},\sum_{i=1}^{m} p_{i2},\ldots,\sum_{i=1}^{m} p_{in}\right)\end{aligned}$$

(*H*4) *Additivity*:

$$H_M(p_1g_1, p_1g_2, \ldots, p_1g_n, p_2g_1, p_2g_2, \ldots,$$
$$p_2g_n, \ldots, p_mg_1, p_mg_2, \ldots, p_mg_n)$$
$$= H_M(p_1, p_2, \ldots, p_m) + H_M(g_1, g_2, \ldots, g_n)$$

for any two probability distributions $(p_1, p_2, \ldots, p_m) \in P$ and $(g_1, g_2, \ldots, g_n) \in P$ of independent sets of events.

(*H*5) *Normalization*:

$$H_M(\tfrac{1}{2}, \tfrac{1}{2}) = 1.$$

(*H*6) *Continuity*: function H_M is continuous in all its arguments $p_1, p_2, \ldots, p_n$. This requirement is often replaced by a weaker requirement of *continuity at* 0:

$$\lim_{p \to 0+} H_M(p, 1-p) = 0.$$

The only entropy which satisfies all these requirements is the Shannon entropy. Each of the other entropies violates at least one of them. As such, the previously mentioned classes of entropies are generalizations of the Shannon entropy in various ways. They are most meaningful for α, β, $R > 0$ since they violate the smallest number of the properties in this range of their parameters: the Rényi entropies violate only the subadditivity property; Daróczy's entropies violate the additivity property; the *R*-norm entropies violate both subadditivity and additivity (they satisfy a weaker property of pseudo-additivity).[7] More details can be found elsewhere[1,7,21,24].

The aim of this paper is to propose an alternative theory of information, one that is developed within the framework of possibility theory[22,33,46–49] rather than probability theory.[30] Such a theory is fundamentally different from the current information theory based on the Shannon entropy and its various generalizations. At the core of the theory is a new conception of information which may conveniently be referred to as *possibilistic information*. It seems that this new conception of information is considerably more meaningful in some situations than the currently prevailing conceptions of uncertainty and information, which are based on the notion of the probability measure.

It is assumed that the possibilistic information is measured in a similar manner as is its probabilistic counterpart, by the amount of uncertainty associated with a finite set of alternative outcomes. The uncertainty with respect to an outcome, characterized by a degree of possibility, is in this sense considered as numerically equal to the information obtained when the outcome actually occurs. Keeping in mind this one-to-one correspondence between information and uncertainty, we may conveniently develop our framework for possibilistic information theory in terms of the underlying notion of *possibilistic uncertainty*.

2 HARTLEY'S MEASURE OF INFORMATION

It is generally recognized that the first measure of information was introduced by Hartley in 1928.[25] He defined the information $I(n)$ necessary to characterize an element of a finite set X with n elements as

$$I(n) = \log_2 n.\dagger$$

This measure is frequently given one of two probabilistic interpretations. In the first, it is viewed as a special entropy, usually called the Hartley entropy, which in a given set of outcomes distinguishes only between zero and nonzero probabilities and which, otherwise, is totally insensitive to the actual values of the probabilities. In this interpretation, the Hartley entropy coincides with the Rényi entropy of degree 0, $H_{\alpha=0}$. It satisfies all the properties mentioned in Section 1 (symmetry, expansibility, etc.) except continuity. It satisfies instead a property of *insensitivity*, defined as

$$H_{\alpha=0}(p_1, 1-p_1) = H_{\alpha=0}(p_2, 1-p_2)$$

for at least one pair p_1, p_2 such that $0 < p_1 < p_2 \leq \frac{1}{2}$. It is also known that every entropy which is symmetric, expansible, subadditive and additive can be expressed as a linear combination of the Shannon and Hartley entropies.[2]

†When the logarithm base 2 is used, the information is expressed in bits.

The second probabilistic interpretation of the Hartley measure $I(n)$ relates it to the Shannon entropy. The Hartley measure becomes equal to the Shannon entropy under the assumption that all elements of the set X with n elements are equally probable.

Although the probabilistic interpretations subsume the Hartley measure under the framework of probabilistic information theory, we believe that such attempts are ill-conceived. Indeed, it is emphasized by Hartley himself that his measure depends solely on the *possibilities* of the individual outcomes and the notion of probability is nowhere even mentioned in his seminal paper.[25] Similar sentiments are expressed by Kolmogorov, who refers to the Hartley measure as "the combinatorial approach to the quantitative definition of information" and writes: "Discussions of information theory do not usually go into this combinatorial approach at any length, but I consider it important to emphasize its logical independence of probabilistic assumptions".[31]

The Hartley measure is rarely mentioned in the current literature on information theory. If it is mentioned at all, it is usually presented in terms of one of the previously mentioned probabilistic interpretations. There are only a few exceptions. One of them is a book by Rényi that contains perhaps the best available discussion of the Hartley measure in which no reference is made to the concept of probability (Ref. 36, pp. 540–546).

Rényi characterizes the Hartley entropy by the following three properties (axioms):

(I1) $I(nm)=I(n)+I(m)$ for $n, m=1,2,\ldots$;

(I2) (I2) $I(n)\leq I(n+1)$;

(I3) $I(2)=1$.

Property (I1) expressed the requirement that the information should be an additive quantity, i.e., if a set X with nm elements can be decomposed into n subsets each with m elements, then we should be able to proceed in two steps in order to characterize an element of X. First, we identify that subset to which the element in question belongs; the required amount of information is $I(n)$. Next we characterize the element within the identified subset with the required amount of information being equal to $I(m)$. The two amounts of information completely characterize an element of the set X.

Property (I2) expresses the requirement that the larger a set, the more information is obtained by the characterization of its elements. Property (I3) is introduced to ensure that the information be measured in bits.

Rényi proves that $I(n)=\log_2 n$ is the only function that satisfies all three properties. Moreover, he proves that the Hartley measure is unique even when property (I1) is weakened to apply only when n and m are relatively prime numbers and when property (I2) is replaced by

$$(I2')\quad \lim_{n\to\infty}(I(n+1)-I(n))=0.$$

The Hartley measure of information is based on the notion of the uncertainty associated with the selection of one element from a set of possible outcomes. In order to use it, possible outcomes must be distinguished, within a given universe of discourse, from those which are not possible. It is thus the *possibility* of an outcome that matters in the Hartley measure, while its probability is totally irrelevant. In this sense, the Hartley measure can be meaningfully generalized only through generalization of the notion of possibility. Such a new avenue opened when Zadeh introduced a theory of possibility in 1978[48] (50 years after this basically possibilistic measure of information was introduced!). Our aim is this paper is to explore this avenue.

3 POSSIBILITY DISTRIBUTIONS AND MEASURES

The notion of possibility distributions associated with a set was proposed by Zadeh[48,49] in his attempt to develop a *theory of possibility* within the general framework of the theory of fuzzy sets. He also introduced the concept of possibility measures, which is a special type of the fuzzy measures proposed by Sugeno.[40]

Let X be a crisp (nonfuzzy) set and let $\mathscr{P}(X)$ denote the power set (crisp) of X. Then, a *fuzzy measure* is a function

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

that has the following properties:

(m1) $m(\varnothing)=0$;

(m2) $A\subseteq B\Rightarrow m(A)\leq m(B)$;

†m can be generalized by replacing $\mathscr{P}(X)$ with a Borel field $\mathscr{B}\subset\mathscr{P}(X)$.

(m3) $A_1 \subseteq A_2 \subseteq \ldots \Rightarrow m\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i \to \infty} m(A_i)$;

(m4) $A_1 \supseteq A_2 \supseteq \ldots \Rightarrow m\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i \to \infty} m(A_i)$.

A *possibility measure* is a function

$$\pi: \mathscr{P}(X) \to [0, 1]$$

that has the following properties:

(π_1) $\pi(\varnothing) = 0$;

($\pi 2$) $A \subseteq B \Rightarrow \pi(A) \leq \pi(B)$;

($\pi 3$) $\pi\left(\bigcup_{i \in I} A_i\right) = \sup_{i \in I} \pi(A_i)$.

Any possibility measure can be uniquely determined by a *possibility distribution function*

$$f: X \to [0, 1]$$

via the formula

$$\pi(A) = \sup_{x \in A} f(x),$$

where $A \subset X$.

It is known[34] that the possibility measure is always a fuzzy measure when set X is finite; when X is not finite, the possibility measure is a fuzzy measure only for some special possibility distribution functions. Hence, it is always assumed in this paper that X is finite. Let $X = \{x_1, x_2, \ldots, x_n\}$ and let

$$f = (\varphi_1, \varphi_2, \ldots, \varphi_n),$$

where $\varphi_i = f(x_i)$, $i \in N_n$, denote the *possibility distribution* associated with function f.† Furthermore, let ${}^n\mathscr{F}$ denote the set of possibility distributions with at least one nonzero element‡ that can be defined on the set X or, generally, on any finite set with n elements. Thus,

$${}^n\mathscr{F} = \{(\varphi_1, \varphi_2, \ldots, \varphi_n) | \varphi_i \in [0, 1],\ i \in N_n,$$
$$(\varphi_1, \varphi_2, \ldots, \varphi_n) \neq (0, 0, \ldots, 0)\}$$

and let

$$\mathscr{F} = \bigcup_{n \in \mathbb{N}} {}^n\mathscr{F}.$$

Let $B(N_n)$ denote the set of all permutations on N_n, i.e., all bijections from N_n to itself. For each $f = (\varphi_1, \varphi_2, \ldots, \varphi_n) \in \mathscr{F}$ and for each $b \in B(N_n)$, let $b[f] = (\varphi_{b(1)}, \varphi_{b(2)}, \ldots, \varphi_{b(n)}) \in \mathscr{F}$.

Before discussing the notion of possibilistic uncertainty, we must first introduce some relevant concepts regarding possibility distributions.

DEFINITION 1 A possibility distribution $f = (\varphi_1, \varphi_2, \ldots, \varphi_n) \in {}^n\mathscr{F}$ is called a normalized possibility distribution if and only if

$$\mathrm{MAX}_i\, \varphi_i = 1.$$

DEFINITION 2 Given a possibility distribution

$$f = (\varphi_1, \varphi_2, \ldots, \varphi_n),$$

let

$$\hat{f} = (\hat{\varphi}_1, \hat{\varphi}_2, \ldots, \hat{\varphi}_n),$$

where $\hat{\varphi}_j = \varphi_{b(j)}$ for some permutation $b \in B(N_n)$ such that $\hat{\varphi}_j \geq \hat{\varphi}_k$ when $j < k$ $(j, k \in N_n)$. Then, $\hat{f}$ is called an *ordered possibility distribution* of f.

DEFINITION 3 Given a possibility distribution f and a permutation $b \in B(N_n)$, if $\hat{f} = b[f]$, then b is called an *ordering permutation* of f.

DEFINITION 4 For each $f = (\varphi_1, \varphi_2, \ldots, \varphi_n) \in \mathscr{F}$ and each $\ell \in [0, 1]$, let

$$c: \mathscr{F} \times [0, 1] \to \mathscr{P}(\mathbb{N})$$

be a function such that

$$c(f, \ell) = \{i \in N_n | \varphi_i \geq \ell\}.$$

†The use of the same symbol for possibility distribution functions and the resulting possibility distributions should not create any confusion since it is stated in each instance which of the two meanings is considered.

‡It is not meaningful to consider a set of alternatives none of which is possible. Although it is reasonable to define uncertainty for such a set to be zero, it is more convenient to exclude this singular case from further considerations as it is of no practical use. Observe that such a set of alternatives is also not acceptable for defining probability distributions and, indeed, the Hartley uncertainty is not defined for it.

This function is called an ℓ*-cut function* and the set $c(f,\ell)$ is called an ℓ*-cut of* f.

DEFINITION 5 Let $f=(\varphi_1,\varphi_2,\ldots,\varphi_n)\in\mathscr{F}$. Then,

$$L_f=\{\ell|(\exists i\in N_n)(\varphi_i=\ell) \text{ or } \ell=0\}$$

is called a *level set* of f.

For each $f=(\varphi_1,\varphi_2,\ldots,\varphi_n)\in\mathscr{F}$, let $L_f=\{\ell_1,\ell_2,\ldots,\ell_r\}$ denote the level set of f, where $\ell_1=0$, $r=|L_f|$ and, $i<j$ implies $\ell_i<\ell_j$. For convenience, let

$$\ell_f=\operatorname*{MAX}_i \varphi_i.$$

Clearly, $\ell_f=\ell_r\in L_f$; moreover, $\ell_f=1$ if and only if f is a normalized possibility distribution.

DEFINITION 6 For every $n\in\mathbb{N}$, let

$${}^1f=({}^1\varphi_1,{}^1\varphi_2,\ldots,{}^1\varphi_n)\in{}^n\mathscr{F}$$

and

$${}^2f=({}^2\varphi_1,{}^2\varphi_2,\ldots,{}^2\varphi_n)\in{}^n\mathscr{F}.$$

be two possibility distributions. Then, ${}^1f$ is called a *subdistribution* of ${}^2f$ if and only if

$${}^1\varphi_i\leqq{}^2\varphi_i \quad \text{for all} \quad i\in N_n;$$

let ${}^1f\leqq{}^2f$ be used to indicate that ${}^1f$ is a subdistribution of ${}^2f$.

The relation "${}^1f$ is a subdistribution of ${}^2f$" introduced in Definition 6 is clearly a partial ordering defined on the set ${}^n\mathscr{F}$. Moreover, $({}^n\mathscr{F},\leqq)$ is a lattice with join and meet defined, respectively, as

$${}^if\vee{}^jf=(\mathrm{MAX}\,[{}^i\varphi_1,{}^j\varphi_1],$$
$$\mathrm{MAX}\,[{}^i\varphi_2,{}^j\varphi_2],\ldots,\mathrm{MAX}\,[{}^i\varphi_n,{}^j\varphi_n])$$
$${}^if\wedge{}^jf=(\mathrm{MIN}\,[{}^i\varphi_1,{}^j\varphi_1],$$
$$\mathrm{MIN}\,[{}^i\varphi_2,{}^j\varphi_2],\ldots,\mathrm{MIN}\,[{}^i\varphi_n,{}^j\varphi_n]).$$

The following lemma establishes an important relation (utilized later in the paper) between the partial ordering introduced in Definition 6 and ℓ-cuts of possibility distributions.

LEMMA 1 *Let* ${}^1f,{}^2f\in{}^n\mathscr{F}$ *and let* $|c({}^1f,\ell)|$, $|c({}^2f,\ell)|$ *denote the cardinalities of* ℓ*-cuts of* ${}^1f$, ${}^2f$, *respectively. Then,*

$$(\forall\ell\in[0,1])(|c({}^1f,\ell)|\leqq|c({}^2f,\ell)|) \tag{1}$$

$$\Leftrightarrow{}^1\hat{f}\leqq{}^2\hat{f} \tag{2}$$

$$\Leftrightarrow(\exists b\in B(N_n))({}^1f\leqq b[{}^2f]) \tag{3}$$

Proof Let ${}^jf=({}^j\varphi_1,{}^j\varphi_2,\ldots,{}^j\varphi_n)$, let ${}^j\hat{f}=({}^j\hat{\varphi}_1,{}^j\hat{\varphi}_2,\ldots,{}^j\hat{\varphi}_n)$, and let ${}^j\hat{f}={}^jb[{}^jf]$, where $j=1,2$.

(i) *Proof of* (1)⇒(2).

Assume (1). Let $\ell_i={}^1\hat{\varphi}_i$ for each $i\in N_n$. Since

$${}^1\varphi_{{}^1b(k)}={}^1\hat{\varphi}_k\geqq{}^1\hat{\varphi}_i=\ell_i \quad \text{for all} \quad k\leqq i,$$

we get

$$c({}^1f,\ell_i)\supseteq\{{}^1b(1),{}^1b(2),\ldots,{}^1b(i)\}.$$

Thus, $|c({}^1f,\ell_i)|\geqq i$ and, from (1), we get $|c({}^2f,\ell_i)|\geqq i$ for all $i\in N_n$. This, together with ${}^2\varphi_{{}^2b(k)}\geqq{}^2\varphi_{{}^2b(j)}$ for $j>k$, implies

$$c({}^2f,\ell_i)\supseteq\{{}^2b(1),{}^2b(2),\ldots,{}^2b(i)\}.$$

Thus, ${}^2\varphi_{{}^2b(k)}\geqq\ell_i$ for all $k\leqq i$. Hence, ${}^2\hat{\varphi}_i={}^2\varphi_{{}^2b(i)}\geqq\ell_i={}^1\hat{\varphi}_i$ for all $i\in N_n$. This, by Definition 6, implies ${}^2\hat{f}\geqq{}^1\hat{f}$.

(ii) *Proof of* (2)⇒(3).

Assume (2). Then, since ${}^j\hat{f}={}^jb[{}^jf]$ $(j=1,2)$, we get ${}^2b[{}^2f]\geqq{}^1b[{}^1f]$ and, clearly,

$${}^1b^{-1}[{}^2b[{}^2f]]\geqq{}^1b^{-1}[{}^1b[{}^1f]]={}^1f,$$

where ${}^1b^{-1}\in B(N_n)$ denotes the inverse of the bijection ${}^1b$. Hence, $b[{}^2f]\geqq{}^1f$, where $b={}^2b\circ{}^1b^{-1}\in B(N_n)$. This implies (3).

(iii) *Proof of* (3)⇒(1).

Assume (3). Let $\ell\in[0,1]$ and let $c({}^1f,\ell)=\{{}^1b(i_1),{}^1b(i_2)\ldots,{}^1b(i_k)\}$ for some $k\in N_n$. Then, by the assumption of (3), we get

$${}^2\varphi_{{}^2b(i_j)}\geqq{}^1\varphi_{{}^1b(i_j)}\geqq\ell$$

for $j=1,2,\ldots,k$. Hence,

$$c({}^2f,\ell)\supseteq\{{}^2b(i_1),{}^2b(i_2),\ldots,{}^2b(i_k)\}$$

and, consequently, $|c({}^2f,\ell)|\geqq k=|c({}^1f,\ell)|$.

From (i), (ii), and (iii) we may conclude that (1)⇔(2)⇔(3), and the lemma is proved. Q.E.D.

4 CHARACTERIZATION OF POSSIBILISTIC UNCERTAINTY

Let

$$V: \mathscr{F} \to [0, \infty)$$

for each $n \in \mathbb{N}$. This function represents a general characterization of a measure of uncertainty for possibility distributions in set $\mathscr{F}$. However, to qualify as an acceptable measure of uncertainty, function V must possess certain properties which we intuitively associate with such a measure. If we want to derive a possibilistic counterpart of the Shannon measure of uncertainty and, at the same time, a generalization of the Hartley measure of uncertainty, we must require that function V satisfy all properties of the Shannon measure for which possibilistic counterparts are meaningful, as well as appropriate generalizations of the properties of the Hartley measure.

Inspecting properties $(H1)$–$(H6)$ of the Shannon measure, defined in Section 1, it is clear that possibilistic counterparts are meaningful for all of them. Two properties of the Hartley uncertainty have counterparts among the properties of the Shannon uncertainty: the additivity $(I1)$ and normalization $(I3)$. In addition, the Hartley uncertainty satisfies a property for which there is no meaningful probabilistic counterpart: Property $(I2)$, which can best be described as *monotonicity*. Since the possibilistic uncertainty is supposed to subsume the Hartley uncertainty as a special case, we have to formulate a property of general monotonicity for the possibilistic measure in such a way that it becomes identical with $(I2)$ for crisp possibility distributions.

It follows from the previous discussion that a desirable possibilistic measure of uncertainty (a possibilistic analog of the Shannon uncertainty) is required to satisfy possibilistic counterparts of symmetry, expansibility, subadditivity, additivity, normalization and continuity, as well as a generalization of monotonicity $(I2)$. These properties are defined as follows:

$(V1)$ Symmetry:

$$V(\varphi_1, \varphi_2, \ldots, \varphi_n) = V(\varphi_{b(1)}, \varphi_{b(2)}, \ldots, \varphi_{b(n)})$$

for all possibility distributions

$f = (\varphi_1, \varphi_2, \ldots, \varphi_n) \in \mathscr{F}$ and all permutations $b \in B(N_n)$.

$(V2)$ *Expansibility*:

$$V(\varphi_1, \varphi_2, \ldots, \varphi_n, 0) = V(\varphi_1, \varphi_2, \ldots, \varphi_n)$$

for all $f = (\varphi_1, \varphi_2, \ldots, \varphi_n) \in \mathscr{F}$.

$(V3)$ *Subadditivity*: Let

$$f = (\varphi_{11}, \varphi_{12}, \ldots, \varphi_{1n}, \varphi_{21}, \varphi_{22}, \ldots, \varphi_{2n}, \ldots, \varphi_{m1}, \varphi_{m2}, \ldots, \varphi_{mn}) \in F,$$

$${}^1f = ({}^1\varphi_1, {}^1\varphi_2, \ldots, {}^1\varphi_m) \in \mathscr{F}$$

such that ${}^1\varphi_i = \operatorname*{MAX}_j \varphi_{ij}$, and

$${}^2f = ({}^2\varphi_1, {}^2\varphi_2, \ldots, {}^2\varphi_n) \in \mathscr{F}$$

such that ${}^2\varphi_j = \operatorname*{MAX}_i \varphi_{ij}$.

Then, $V(f) \leq V({}^1f) + V({}^2f)$.

$(V4)$ *Additivity*: In $(V3)$, if ${}^1f$ and ${}^2f$ are non-interactive, i.e., $\varphi_{ij} = \text{MIN}[{}^1\varphi_i, {}^2\varphi_j]$ for all $i \in N_m$ and all $j \in N_n$, then the equality holds, i.e.,

$$V(f) = V({}^1f) + V({}^2f).$$

$(V5)$ *Normalization*: $V(1, 1) = 1$.

$(V6)$ *Continuity*: function V is continuous in all its arguments.

$(V7)$ *Monotonicity*: Let ${}^1f, {}^2f \in {}^n\mathscr{F}$. Then, $V({}^1f) \leq V({}^2f)$ if ${}^1f \leq {}^2f$ and

$$\operatorname*{MAX}_i {}^1\varphi_i = \operatorname*{MAX}_i {}^2\varphi_i.$$

Properties $(V1)$, $(V2)$, $(V3)$, $(V5)$, $(V6)$ are clearly counterparts of properties $(H1)$, $(H2)$, $(H3)$, $(H5)$, $(H6)$, respectively. Properties $(V4)$ and $(V7)$ need some discussion.

Additivity $(V4)$ is obviously introduced as a counterpart of additivity $(H4)$ for the Shannon measure. Observe that the probabilistic additivity is required only when the two probability distributions are associated with independent sets of events. The possibilistic additivity, on the other hand, is required only for possibility distributions of non-interactive sets of events. Let us explain why this formulation of $(V4)$ is necessary.

It is well known that the concepts of non-interaction and independence are equivalent within probability theory.[26]† Hence, ($H4$) can be formulated in terms of independence or in terms of non-interaction (or both). It is not obvious, however, that these concepts are equivalent when defined in terms of possibility theory. Two views regarding this issue have been expressed in the literature by Hisdal[26] and Nguyen.[32]

Given two sets of events X and Y (finite for our purpose) with possibility distributions $f(x)$ and $f(y)$ ($x \in X$, $y \in Y$), respectively, they are called non-interactive if and only if

$$f(x, y) = \text{MIN}\,[f(x), f(y)] \tag{4}$$

for all $x \in X$ and $y \in Y$, where $f(x, y)$ denotes the joint possibilities of x and y. The sets are called independent if and only if

$$f(x|y) = f(x) \tag{5}$$

and

$$f(y|x) = f(y) \tag{6}$$

for all $x \in X$ and $y \in Y$, where $f(x|y)$ and $f(y|x)$ denote the conditional possibilities of x given y and y given x, respectively.

Hisdal[26] argues that the equations

$$f(x, y) = \text{MIN}\,[f(y), f(x|y)] \tag{7}$$

$$f(x\, y) = \text{MIN}\,[f(x), f(y|x)] \tag{8}$$

must be satisfied for any two sets of events characterized by some possibility distributions $f(x)$, $f(y)$. Then, it is clear that (5) and (7) as well as (6) and (8) imply (4). Hence, the independence of possibilistic events implies their non-interaction. However, the converse is not true. Indeed, from (4) and (7) we obtain

$$f(x|y) = \begin{cases} f(x) & \text{if } f(x) < f(y) \\ [f(x), 1] & \text{if } f(x) \geqq f(y) \end{cases} \tag{9}$$

(and similarly for the other conditional possibility). Hence, independence does not imply non-interaction.

Nguyen[32] takes a radically different approach to the meaning of the conditional possibilities. He defines "normalized" conditional possibilities in such a way that, by analogy with probability theory, possibilistic noninteraction is required to be equivalent to possibilistic independence. This requirement leads to the formula

$$f(x|y) = \begin{cases} f(x, y) & \text{if } f(x) \leqq f(y) \\ f(x, y) \cdot \dfrac{f(x)}{f(y)} & \text{if } f(x) > f(y), \end{cases} \tag{10}$$

where $f(x)/f(y)$ is a normalization factor.

In our opinion, the normalization requirement proposed for conditional possibilities by Nguyen is not well justified. In any case, the meaning of conditional possibilities and, consequently, the relationship between possibilistic non-interaction and independence remain somewhat controversial. To avoid this controversy, we require in our characterization of the possibilistic uncertainty, that the additivity property ($V4$) be satisfied for non-interactive sets of events. According to Hisdal, possibilistic independence implies non-interaction; according to Nguyen, the two concepts are equivalent. Hence, regardless of which of the two views one decides to accept, ($V4$) implies the following weaker property:

($V4'$) The equality in ($V3$) hold if ${}^1f$ and ${}^2f$ are independent.

Let us demonstrate now that properties ($V4$), ($V5$), ($V7$) are natural generalizations of properties ($I1$), ($I3$), ($I2$), respectively, which are used by Rényi to characterize the Hartley uncertainty.[36] To make this comparison possible, let

$${}^n\mathscr{F}_c = \{f = (\varphi_1, \varphi_2, \ldots, \varphi_n) \mid \varphi_i \in \{0, 1\},$$
$$i \in N_n, (\varphi_1, \varphi_2, \ldots, \varphi_n) \neq (0, 0, \ldots, 0)\}$$

and let

$$\mathscr{F}_c = \bigcup_{n \in \mathbb{N}} {}^n\mathscr{F}_c$$

†Let $x \in X$ and $y \in Y$, where X and Y are some finite sets of events. Probabilistic non-interaction is defined by $p(x, y) = p(x) \cdot p(y)$ for all $x \in X$ and all $y \in Y$, where $p(x, y)$ denotes the joint probabilities. Probabilistic independence is defined by $p(x|y) = p(x)$ and $p(y|x) = p(y)$, where $p(x|y)$ denotes the conditional probability of x given y and $p(y|x)$ denotes the other conditional probability.

denote the set of all crisp possibility distributions defined on a set with n elements. Due to the requirements of symmetry ($V1$) and expansibility ($V2$), the value of the uncertainty $V(f)$ for any $f \in \mathscr{F}_c$ must depend solely on the number of possible outcomes, which is the same as the cardinality of $c(f, 1)$ (1-cut of f), i.e., for every $f \in \mathscr{F}_c$,

$$V(f) = I'(n), \tag{11}$$

where $I': \mathbb{N} \rightarrow [0, 1]$ and $n = |c(f, 1)|$. Similarly, the Hartley uncertainty is defined in terms of the number of alternative outcomes, i.e., the cardinality of a finite set. Consequently, $V(f)$ is comparable with the Hartley uncertainty for crisp possibility distributions ($f \in \mathscr{F}_c$). Applying (11) to ($V4$), it can easily be shown that ($V4$) and ($I1$) are equivalent when ${}^1f$ and ${}^2f$ in ($V4$) are crisp possibility distributions ($c({}^1f, 1) = m$, $c({}^2f, 1) = n$).

It follows immediately from (11) that the normalization property ($V5$), by which it is resolved that the uncertainty be measured in bits, coincides with ($I3$). It is also a trivial consequence of (11) that ($V7$) is equivalent to ($I2$) when ${}^1f$ and ${}^2f$ in ($V7$) are crisp possibility distributions (${}^1f, {}^2f \in \mathscr{F}_c$).

The Rényi properties ($I1$)–($I3$) can thus be derived from properties ($V1$)–($V7$) when the possibility distributions involved are crisp. From this fact, Eq. (11), and the uniqueness of the Hartley uncertainty,[36] we may now conclude that

$$V(f) = \log_2 |c(f, 1)|$$

for all $f \in \mathscr{F}_c$. Hence, V coincides with the Hartley uncertainty within the subset $\mathscr{F}_c$ of $\mathscr{F}$.

If we appeal to our intuition for some additional desirable properties of the possibilistic measure of uncertainty, the following are likely to emerge and may be added to the list of properties ($V1$)–($V7$):

($V8$) For all $f = (\varphi_1, \varphi_2, \ldots, \varphi_n) \in \mathscr{F}$, $V(f) = 0$ if and only if $\varphi_i \neq 0$ for exactly one $i \in N_n$.

($V9$) For all $f = (\varphi_1, \varphi_2, \ldots, \varphi_n) \in \mathscr{F}$, $V(f)$ attains its maximum within ${}^n\mathscr{F}$, if and only if φ_i has the same value for all $i \in N_n$; the maximum of $V(f)$ within ${}^n\mathscr{F}$ is equal to $V(1, 1, \ldots, 1) = \log_2 n$.

($V10$) For all $f = (\varphi_1, \varphi_2, \ldots, \varphi_n) \in \mathscr{F}$, $V(f)$ decreases if $V(f) \neq 0$ and only one maximum elements of f, say element φ_{i_0} such that $\varphi_{i_0} = \mathrm{MAX}_i \, \varphi_i$, increases.

Properties ($V8$) and ($V9$), which may conveniently be referred to as *extremal properties* of V, are, indeed, highly desirable. Property ($V10$), which is applicable only to non-normalized possibility distributions (Def. 1), also seems intuitively quite desirable.

5 PROPOSED MEASURE OF POSSIBILISTIC UNCERTAINTY

Our aim in this section is to determine a function that satisfies all the properties ($V1$)–($V7$) required for a measure of possibilistic uncertainty and, preferably, also the additional properties ($V8$)–($V10$). Let us denote such a function by U, i.e.,

$$U: \mathscr{F} \rightarrow [0, \infty).$$

In order to determine an intuitively meaningful function U, let us consider the following.

Let $f = (\varphi_1, \varphi_2, \ldots, \varphi_n) \in \mathscr{F}$ be given and let $L_f = \{\ell_1 = 0, \ell_2, \ldots, \ell_r = \ell_f\}$ be the level set of f (Def. 5). For each pair ℓ_k, $\ell_{k+1} \in L_f$ ($l = 1, 2, \ldots, r-1$), we can define a dichotomy of the set of outcomes into possible and impossible outcomes by using the following rule: an outcome is possible if its associated possibility degree is not smaller than ℓ_{k+1}; it is not possible if it is not larger than ℓ_k. It is reasonable to view this dichotomy as an approximation of f with respect to levels ℓ_k, ℓ_{k+1} by a crisp possibility distribution

$$f_{\ell_k, \ell_{k+1}} = (\varphi'_1, \varphi'_2, \ldots, \varphi'_n)$$

such that

$$\varphi'_i = \begin{cases} 1 & \text{if } \varphi_i \geqq \ell_{k+1} \text{ (i.e., } i \in c(f, \ell_{k+1})) \\ 0 & \text{if } \varphi_i \leqq \ell_k \text{ (i.e., } i \notin c(f, \ell_{k+1})) \end{cases}$$

Let $f_{\ell_k, \ell_{k+1}}$ denote (ℓ_k, ℓ_{k+1})–crisp approximation of f. As shown previously, U must coincide with the Hartley uncertainty for crisp possibility distributions. Hence,

$$U(f_{\ell_k, \ell_{k+1}}) = \log_2 |c(f, \ell_{k+1})|$$

Taking the mean of uncertainties for all (ℓ_k, ℓ_{k+1})-crisp approximations of f with weights representing the gaps between the edge values of

ℓ_k and ℓ_{k+1}, i.e., $\ell_{k+1}-\ell_k$, we obtain

$$U(f)=\frac{1}{\ell_f}\sum_{k=1}^{r-1}(\ell_{k+1}-\ell_k)\log_2|c(f,\ell_{k+1})| \quad (12)$$

Formula (12) defines the measure of possibilistic uncertainty which is proposed in this paper. Let us refer to this measure as the U-uncertainty. Observe that (12) can also be written as

$$U(f)=\frac{1}{\ell_f}\int_0^{\ell_f}\log_2|c(f,\ell)|\,d\ell. \quad (13)$$

Although the derivation of U-uncertainty is based on reasonable considerations, we still must prove that it satisfies all the required properties. First, let us prove a useful lemma.

LEMMA 2 *Let* ${}^1f,\ {}^2f\in{}^n\mathscr{F}$. *Then,* $(V1)$ *and* $(V7)$, *taken together, are equivalent to*

$$(V7')\quad V({}^1f)\leqq V({}^2f)\quad \textit{if}\quad \operatorname*{MAX}_i{}^1\varphi_i=\operatorname*{MAX}_i{}^2\varphi_i$$

$$\textit{and}\ (\exists b\in B(N_n))({}^1f\leqq b[{}^2f])$$

as well as to

$$(V7'')\quad V({}^1f)\leqq V({}^2f)\ \textit{if}\ \operatorname*{MAX}_i{}^1\varphi_i=\operatorname*{MAX}_i{}^2\varphi_i$$

$$\textit{and}\ (\forall\ell\in[0,1])(|c({}^1f,\ell)|\leqq|c({}^2f,\ell)|).$$

Proof Let ${}^1f,\ {}^2f\in{}^n\mathscr{F}$ and let $\mathrm{MAX}_i{}^1\varphi_i=\mathrm{MAX}_i{}^2\varphi_i$.

(i) Assume that the function $V:\mathscr{F}\to[0,\infty)$ under consideration satisfies the properties $(V1)$ and $(V7)$. Assume further that the statement

$$(\exists b\in B(N_n))({}^1f\leqq b[{}^2f])$$

is true. Then, by $(V7)$, we obtain

$$V({}^1f)\leqq V(b[{}^2f]),$$

and, from $(V1)$, we obtain

$$V(b[{}^2f])=V({}^2f).$$

Hence, $V({}^1f)\leqq V({}^2f)$.

(ii) Assume that V satisfies $(V7')$ and ${}^1f\leqq{}^2f$. Then, clearly ${}^2f=b[{}^2f]$, where $b\in B(N_n)$ is the identity function on N_n, and according to $(V7')$, $V({}^1f)\leqq V({}^2f)$. This implies $(V7)$. Assume now that the statement

$$(\exists b\in B(N_n))({}^1f=b[{}^2f])$$

is true. Then, ${}^1f\leqq b[{}^2f]$ and ${}^1f\geqq b[{}^2f]$ and from $(V7')$ we get

$$V({}^1f)\leqq V(b[{}^2f])$$

and

$$V({}^1f)\geqq V(b[{}^2f]).$$

Hence,

$$V({}^1f)=V(b[{}^2f]),$$

which implies $(V1)$.

It follows from (i) and (ii) that $(V1)$ and $(V7)$ are equivalent to $(V7')$. It follows directly from Lemma 1 that $(V7')$ is equivalent to $(V7'')$. Q.E.D.

Now we come to the main issue of this paper: a formal justification of the proposed U-measure as a meaningful measure of possibilistic uncertainty.

THEOREM 1 *Function* U *defined by* (12) *or, alternatively,* (13) *satisfies all the properties* $(V1)$–$(V10)$.

Proof (i) $(V1)$ and $(V7)$: Let us consider, for any $n\in\mathbb{N}$, ${}^1f,{}^2f\in{}^n\mathscr{F}$ such that

$$|c({}^1f,\ell)|\leqq|c({}^2f,\ell)|\ \text{for all}\ \ell\in[0,1].$$

Then, clearly,

$$\log_2|c({}^1f,\ell)|\leqq|c({}^2f,\ell)|\ \text{for all}\ \ell\in[0,1]$$

and, consequently, $U({}^1f)\leqq U({}^2f)$. Hence, U satisfies $(V7'')$ and, by Lemma 2, it also satisfies $(V1)$ and $(V7)$.

(ii) $(V2)$: Let $f=(\varphi_1,\ \varphi_2,\ldots,\ \varphi_n)\in\mathscr{F}$ and $f'\in(\varphi_1,\ \varphi_2,\ldots,\ \varphi_n,\ 0)\in\mathscr{F}$. Then, $L_f=L_{f'}$ and $c(f,\ell)=c(f',\ell)$ for all $\ell\neq0$, but $c(f,0)$ is not included in formula (12). Hence, U satisfies $(V2)$.

(iii) $(V3)$: Let, f, ${}^1f$ and ${}^2f$ be defined as in $(V3)$. Since for each

$$\ell\in[0,1]\varphi_{ij}\geqq\ell\ \text{implies}\ {}^1\varphi_i=\operatorname*{MAX}_j\varphi_{ij}\geqq\ell$$

and

$$^2\varphi_j = \underset{i}{\mathrm{MAX}}\, \varphi_{ij} \geqq \ell,$$

we get

$$\begin{aligned} c(f,\ell) &= \{(i,j) \in N_m \times N_n \mid \varphi_{ij} \geqq \ell\} \\ &\subseteq \{(i,j) \in N_m \times N_n \mid {}^1\varphi_i \geqq \ell \text{ and } {}^2\varphi_j \geqq \ell\} \\ &= \{i \in N_m \mid {}^1\varphi_i \geqq \ell\} \times \{j \in N_n \mid {}^1\varphi_j \geqq \ell\} \\ &= c({}^1f,\ell) \times c({}^2f,\ell). \end{aligned}$$

Hence,

$$|c(f,\ell)| \leqq |c({}^1f,\ell)| \times |c({}^2f,\ell)| \quad \text{for all } \ell \in [0,1].$$

By definition,

$$\ell_f = \underset{i,j}{\mathrm{MAX}}\, \varphi_{ij} = \underset{i}{\mathrm{MAX}}\left(\underset{j}{\mathrm{MAX}}\, \varphi_{ij}\right) = \underset{i}{\mathrm{MAX}}\, {}^1\varphi_i = \ell_{{}^1f}.$$

Similarly, $\ell_f = \ell_{{}^2f}$. Hence,

$$\begin{aligned} U(f) &= \frac{1}{\ell_f}\int_0^{\ell_f} \log_2 |c(f,l)|\, d\ell \\ &\leqq \frac{1}{\ell_f}\int_0^{\ell_f} \log_2 |c({}^1f,\ell)| \times |c({}^2f,\ell)|\, d\ell \\ &= \frac{1}{\ell_f}\int_0^{\ell_f} \log_2 |c({}^1f,\ell)|\, d\ell + \frac{1}{\ell_f}\int^{\ell_f} \log_2 |c({}^2f,\ell)|\, d\ell \\ &= U({}^1f) + U({}^2f). \end{aligned}$$

This concludes the proof that U satisfies $(V3)$.

(iv) $(V4)$: Let f, ${}^1f$ and ${}^2f$ be defined as in $(V4)$. Then, clearly, $\varphi_{ij} \geqq \ell \Leftrightarrow {}^1\varphi_i \geqq \ell$ and ${}^2\varphi_j \geqq \ell$ for all $\ell \in [0,1]$. Hence, as in (iii), we get $c(f, \ell) = c({}^1f, \ell) \times c({}^2f, \ell)$ and, consequently, $|c(f, \ell)| = |c({}^1f, \ell)| \times |c({}^2f, \ell)|$ for all $\ell \in [0,1]$. Following the scheme of reasoning in (iii), we can show that $\ell_f = \ell_{1_f} = \ell_{2_f}$ and, eventually, $U(f) = U({}^1f) + U({}^2f)$. Therefore, U satisfies $(V4)$.

(v) $(V5)$: $U(1,1) = \log_2 2 = 1$.

(vi) $(V6)$: Let $f = (\varphi_1, \varphi_2, \ldots, \varphi_n) \in \mathscr{F}$, let $\alpha \in N_n$ be a particular integer, and let $L_f = \{\ell_1, \ell_2, \ldots, \ell_r\}$, where $0 = \ell_1 < \ell_2 < \ldots < \ell_r (= \ell_f)$. For the sake of clarity, let us distinguish three cases.

Case I Assume $\varphi_\alpha = \ell_\beta$ for some β such that $2 \leqq \beta \leqq r-1$.

(A) Let $\Delta\ell$ denote a real number such that $0 < \Delta\ell < \ell_{\beta+1} - \ell_\beta$ and let $f' = (\varphi'_1, \varphi'_2, \ldots, \varphi'_r)$ denote a possibility distribution such that $\varphi'_i = \varphi_i$ for all $i \neq \alpha$ and $\varphi'_\alpha = \varphi_\alpha + \Delta\ell$. Then, for all $\ell \in [0,1]$, we get

$$\begin{aligned} c(f',\ell) - \{\alpha\} &= \{i \in N_n \mid \varphi'_1 \geqq \ell\} - \{\alpha\} \\ &= \{i \in N_n \mid i \neq \alpha, \varphi'_i \geqq \ell\} \\ &= \{i \in N_n \mid i \neq \alpha, \varphi_i \geqq \ell\} \\ &\quad \{i \in N_n \mid \varphi_i \geqq \ell\} - \{\alpha\} \\ &= c(f,\ell) - \{\alpha\} \end{aligned}$$

For $\ell \leqq \varphi_\alpha$, we also get $\ell \leqq \varphi'_\alpha$ (since $\varphi'_\alpha \geqq \varphi_\alpha$) and, consequently, both $\alpha \in c(f, \ell)$ and $\alpha \in c(f', \ell)$; hence, $c(f', \ell) = c(f, \ell)$. For $\varphi_\alpha < \ell \leqq \varphi_\alpha + \Delta\ell (= \varphi'_\alpha)$, it is clear that $\alpha \notin c(f, \ell)$ and $\alpha \in c(f', \ell)$; hence, $|c(f', \ell)| = |c(f, \ell)| + 1$. For $\ell > \varphi_\alpha + \ell$, both $\alpha \notin c(f, \ell)$ and $\alpha \notin c(f', \ell)$; hence, again, $c(f', \ell) = c(f, \ell)$. We may conclude now that

$$U(f') - U(f) = \frac{1}{\ell_f}\int_{\varphi_\alpha}^{\varphi_\alpha+\Delta\ell} [\log_2(|c(f,\ell)|+1) - \log_2 |c(f,\ell)|\, d\ell],$$

where $c(f,\ell) = c(f,\ell_\beta)$ for all $\ell \in [\varphi_\alpha, \varphi_\alpha + \Delta\ell]$ (Since $\ell_\beta = \varphi_\alpha$ and $\varphi_\alpha + \Delta\ell < \ell_{\beta+1}$). Hence,

$$\begin{aligned} U(f') - U(f) &= \frac{1}{\ell_f}[\log_2(|c(f, \ell_\beta)|+1) \\ &\quad - \log_2 |c(f, \ell_\beta)|]\Delta\ell. \\ &= \frac{1}{\ell_f}\log_2 \frac{|c(f,\ell_\beta)|+1}{c(f,\ell_\beta)}\Delta\ell. \qquad (*) \end{aligned}$$

Since

$$\frac{1}{\ell_f}\log_2 \frac{|c(f,\ell_\beta)|+1}{|c(f,\ell_\beta)|}$$

is a constant, equation (*) implies that U is continuous at f from the right with respect to its argument φ_α.

(B) Let $\ell_{\beta-1} - \ell_\beta < \Delta\ell < 0$, and let f' be defined in the same way as in (A). Then, following the same reasoning used in (A), we obtain

$$U(f')-U(f)=\frac{1}{\ell_f}\log_2\frac{|c(f,\ell_{\beta-1})|}{|c(f,\ell_{\beta-1})|-1}\Delta\ell, \quad (**)$$

which implies that U is continuous from the left with respect to φ_α.

From (A) and (B), we conclude that U is continuous with respect to φ_α.

Case II Assume $\varphi_\alpha=\ell_r$.

(A) Set $\ell_r<1$. First, let $0<\Delta\ell<1-\ell_r$. Following the same reasoning used in Case I (A), we get

$$|U(f')-U(f)|=\left|\left(\frac{1}{\ell_{f'}}-\frac{1}{\ell_f}\right)\cdot C\right|$$

$$=\frac{C}{\ell_r\cdot(\ell_r+\Delta\ell)}\Delta\ell<\frac{C}{\ell_r^2}\Delta\ell,$$

where $C=\int_0^{\ell_r}\log_2|c(f,\ell)|d\ell$ and ℓ_r^2 are constant. Hence, U is continuous from the right. Next, let $\ell_r-\ell_{r-1}<\Delta\ell<0$. Suppose $|c(f,\ell_r)|=1$. Then, by the same reasoning used in the first part of this case (i.e., Case II (A)), we get

$$|U(f')-U(f)|\leqq\frac{C'}{\ell_f\ell_{r-1}}|\Delta\ell|$$

where, $C'=\int_0^{\ell_{r-1}}\log_2|c(f,\ell)|d\ell$ and $\ell_f\ell_{r-1}$ are constant. Hence, U is continuous from the left. If $|c(f,\ell_r)|\neq 1$, then all the arguments in Case I (B) are valid, and we get (**), which implies that U is continuous from the left. Hence, U is continuous with respect to φ_α.

(B) Set $\ell_r=1$. All the arguments in the second part of (A) are valid and, thus, U is continuous from the left with respect to φ_α.

Case III Assume $\varphi_\alpha=\ell_1=0$. Then, all arguments in Case I (A) are valid and U is thus continuous from the right with respect to φ_α.

It follows from Cases I, II, III that U is continuous at any $f\in\mathscr{F}$ with respect to any of its arguments.

(vii) $(V8)$: Let $f=(\varphi_1,\varphi_2,\ldots,\varphi_n)\in\mathscr{F}$ and let

$$L_f=\{\ell_1,\ell_2,\ldots,\ell_r\}.$$

Then, by the definition of U, we get

$$U(f)=0\Leftrightarrow|c(f,\ell_2)|=1 \text{ and } r=2$$
$$0\Leftrightarrow\varphi_i\neq 0 \text{ for exactly one } i\in N_n.$$

Hence, U satisfies $(V8)$.

(viii) $(V9)$: Let $n\in N$, $f=(\varphi_1,\varphi_2,\ldots,\varphi_n)$ and $L_f=\{\ell_1,\ell_2,\ldots,\ell_r=\ell_f\}$. Since $|c(f,\ell_{i+1})|\leqq n$ for all $i\in N_{r-1}$, we get

$$U(f)=\frac{1}{\ell_r}\sum_{i=1}^{r-1}(\ell_{i+1}-\ell_i)\log_2|c(f,\ell_{i+1})|$$

$$\leqq\frac{1}{\ell_r}\sum_{i=1}^{r-1}((\ell_{i+1}-\ell_i)\log_2 n$$

$$=\frac{1}{\ell_r}\log_2 n\sum_{i=1}^{r-1}(\ell_{i+1}-\ell_i)$$

$$=\log_2 n.$$

Hence, $U(f)\leqq\log_2 n$, where the equality holds if and only if $r=2$ and $|c(f,\ell_2)|=n$, i.e., if and only if $\varphi_i=c$ for some constant $c>0$ and all $i\in N_n$. We can thus conclude that U satisfies $(V9)$.

(ix) $(V10)$: Let us consider $f=(\varphi_1,\varphi_2,\ldots,\varphi_n)\in\mathscr{F}$ such that $U(f)\neq 0$. Let $I_f=\{i\in N_n|\varphi_i\geqq\varphi_j$ for all $j\in N_n\}$. Since $U(f)\neq 0$, it follows from (vii) that $|\{i\in N_n|\varphi_i>0\}|\geqq 2$ and, consequently, $\varphi_i>0$ for all $i\in I_f$.

Let $\lambda\in I_f$. Consider $f^\lambda=(\varphi_1^\lambda,\varphi_2^\lambda,\ldots,\varphi_n^\lambda)\in\mathscr{F}$ such that

$$\varphi_i^\lambda\begin{cases}=\varphi_i & \text{for } i\neq\lambda\\ >\varphi_i & \text{for } i=\lambda\end{cases}$$

Then, clearly,

$$c(f^\lambda,\ell)=\begin{cases}c(f,\ell) & \text{for } 0\leqq\ell\leqq\varphi_\lambda, \text{ and}\\ \{\lambda\} & \text{for } \varphi_\lambda\leqq\ell\leqq\varphi_\lambda^\lambda.\end{cases}$$

Hence,

$$U(f^\lambda)=\frac{1}{\varphi_\lambda^\lambda}\int_0^{\varphi_\lambda^\lambda}\log|c(f^\lambda,\ell)|d\ell$$

$$=\frac{1}{\varphi_\lambda^\lambda}\left(\int_0^{\varphi_\lambda}\log|c(f^\lambda,\ell)|d\ell+\int_{\varphi_\lambda}^{\varphi_\lambda^\lambda}\log|c(f^\lambda,\ell)|d\ell\right)$$

$$=\frac{1}{\varphi_\lambda^\lambda}\left(\int_0^{\varphi_\lambda}\log|c(f^\lambda,\ell)|d\ell+\int_{\varphi_\lambda}^{\varphi_\lambda^\lambda}\log 1\,d\ell\right)$$

$$=\frac{1}{\varphi_\lambda^\lambda}(\varphi_\lambda \cdot U(f))$$

$$=\frac{\varphi_\lambda}{\varphi_\lambda^\lambda}U(f).$$

Since

$$\varphi_\lambda < \varphi_\lambda^\lambda, \frac{\varphi_\lambda}{\varphi_\lambda^\lambda} < 1 \quad \text{and} \quad U(f) \neq 0,$$

we get $U(f^\lambda) < U(f)$, which proves that U satisfies ($V10$).

The theorem is proved by (i)—(vix). Q.E.D.

We can now compare function U with another function, say a: $\mathscr{F} \to [0, \infty)$ which was proposed as a measure of possibilistic uncertainty (non-specificity, ambiguity, anxiety) by Yager[44,45] and defined as

$$a(f) = 1 - \sum_{i=1}^{r-1} \frac{\ell_{i+1} - \ell_i}{|c(f, \ell_{i+1})|}$$

or, alternatively,

$$a(f) = 1 - \int_0^{\ell_f} \frac{1}{|c(f, \ell)|} d\ell. \tag{14}$$

It can easily be verified that this measure satisfies only some of the properties ($V1$)–($V10$) required for a measure of possibilistic uncertainty, namely, properties ($V1$), ($V2$), ($V6$), ($V7$), ($V9$) and ($V10$). It does not satisfy properties ($V3$), ($V4$), ($V5$), and ($V8$). Among these properties, ($V5$) is not particularly important, since it only defines the units in which the uncertainty is to be measured, and properties ($V8$), ($V10$) are not relevant if we accept only normalized possibility distributions. However, subadditivity ($V3$) and additivity ($V4$) are essential properties for any measure of uncertainty,[1,2,24,35] especially for a measure which is supposed to be a possibilistic counterpart of the Shannon uncertainty. We also observe that the Yager measure does not coincide with the well established Hartley uncertainty for crisp possibility distributions.

6 CONDITIONAL POSSIBILISTIC UNCERTAINTY

Let

$$f: X \times Y \to [0, 1]$$

be a possibility distribution function, where $X = \{x_1, x_2, \ldots, x_n\}$ and $Y = \{y_1, y_2, \ldots, y_m\}$. Let

$$f = (\varphi_{11}, \varphi_{12}, \ldots, \varphi_{1m}, \varphi_{21}, \varphi_{22}, \ldots, \varphi_{2m}, \ldots, \varphi_{n1}, \varphi_{n2}, \ldots, \varphi_{nm}),$$

where $\varphi_{ij} = f(x_i, y_j)$ denote the possibility distribution associated with function f; it is called the *joint possibility distribution* of X and Y.

Let

$${}^Xf: X \to [0, 1]$$

and

$${}^Yf: Y \to [0, 1]$$

be possibility distribution functions such that

$${}^Xf(x_i) = \operatorname*{MAX}_j f(x_i, y_j)$$

and

$${}^Xf(y_j) = \operatorname*{MAX}_i f(x_i, y_j),$$

respectively; ${}^Xf$ and ${}^Yf$ are called *marginal possibility distribution functions* of f on X and Y, respectively.

In terms of the marginal possibility distribution functions ${}^Xf$, ${}^Yf$; we can define two *marginal possibility distributions*

$${}^Xf = ({}^X\varphi_1, {}^X\varphi_2, \ldots, {}^X\varphi_n)$$

and

$${}^Yf = ({}^Y\varphi_1, {}^Y\varphi_2, \ldots, {}^Y\varphi_n)$$

where ${}^X\varphi_i = {}^Xf(x_i)$, ${}^Y\varphi_j = {}^Yf(y_j)$ and, obviously,

$${}^X\varphi_i = \operatorname*{MAX}_j f(x_i, y_j) = \operatorname*{MAX}_j \varphi_{ij} \tag{15}$$

$${}^Y\varphi_j = \operatorname*{MAX}_i f(x_i, y_j) = \operatorname*{MAX}_i \varphi_{ij} \tag{16}$$

Let f_c denote a crisp possibility distribution on the set $X \times Y$, i.e., a distribution in which $\varphi_{ij} \in \{0,1\}$ for all $i \in N_n$ and $j \in N_m$. Let ${}^Xf_c$ and ${}^Yf_c$ denote the marginal crisp possibility distributions.

For convenience, let us introduce the following notational convention. Given a possibility distribution f defined on $X \times Y$, let $U(X, Y)$, $U(X)$, $U(Y)$ denote the U-uncertainty (defined by (12)) of f, ${}^Xf$, ${}^Yf$, respectively. Similarly, let $U_c(X, Y)$, $U_c(X)$, $U_c(Y)$ denote the U-uncertainty of crisp distributions f_c, ${}^Xf_c$, ${}^Yf_c$, respectively. Furthermore, let $U(Y|X)$, $U(X|Y)$ denote conditional U-uncertainties of ${}^Yf$ given ${}^Xf$ and ${}^Xf$ given ${}^Yf$, respectively ($U_c(Y|X)$ and $U_c(X|Y)$ for crisp distributions). The purpose of this section is to clarify the meaning of the conditional U-uncertainties.

It is easy to see that the ratio

$$\frac{|c(f_c, 1)|}{|c({}^Xf_c, 1)|}$$

represents the average number of elements of Y that occur (i.e., ${}^Y\varphi_j = 1$) under the condition that an element of X occurs (i.e., ${}^X\varphi_i = 1$). The binary logarithm of this ratio, then, represents the conditional U-uncertainty (crisp) of ${}^Yf_c$ given ${}^Xf_c$. Hence.

$$U_c(Y|X) = \log_2 \frac{|c(f_c, 1)|}{|c({}^Xf_c, 1)|}$$

$$= U_c(X, Y) - U_c(X).$$

Given a general possibility distribution f defined on $X \times Y$, we can see that, as in the case of a crisp distribution f_c, the ratio

$$\frac{|c(f, \ell)|}{|c({}^Xf, \ell)|}$$

represents for each level $\ell \in [0, \ell_f]$ the average number of elements of Y that occur (i.e., ${}^Y\varphi_j \geqq \ell$) under the condition that an element of X occurs (i.e., ${}^X\varphi_i \geqq \ell$).

Hence, the binary logarithm of this ratio represents the conditional U-uncertainty of ${}^Yf$ given ${}^Xf$ with respect to level ℓ (i.e., for ℓ-cuts, which are crisp sets). The overall U-uncertainty of ${}^Yf$ given ${}^Xf$ should express the mean of this ℓ-dependent uncertainty for all $\ell \in [0, \ell_f]$, i.e.,

$$U(Y|X) = \frac{1}{\ell_f} \int_0^{\ell_f} \log_2 \frac{|c(f, \ell)|}{|c({}^Xf, \ell)|} d\ell$$

$$= \frac{1}{\ell_f} \int_0^{\ell_f} \log_2 |c(f, \ell)|\, d\ell$$

$$- \frac{1}{\ell_f} \int_0^{\ell_f} \log_2 |c({}^Xf, \ell)|\, d\ell$$

$$= U(X, Y) - U(X). \tag{17}$$

In a similar way, we can derive that

$$U(X|Y) = U(X, Y) - U(Y). \tag{18}$$

Equations (17) and (18) are significant in two respects. First, they have exactly the same form as their counterparts for the Shannon entropy and, second, they permit computation of the two conditional U-uncertainties without using conditional possibilities at all, thus avoiding the controversy that currently surrounds them.[26,32]

7 CONCLUSIONS

A measure of uncertainty for possibility theory[22,33,46–49] is introduced in this paper. The measure is referred to as the U-uncertainty and is defined by Eq. (12) or, alternatively, Eq. (13). It can be viewed as a possibilistic counterpart of the Shannon entropy[1,37] and, at the same time, a generalization of the Hartley uncertainty.[25]

The U-uncertainty satisfies, in agreement with the Shannon entropy, the possibilistic counterparts ($V1$)–($V6$) of the probabilistic properties of symmetry, expansibility subadditivity, additivity, normalization, and continuity, respectively.[1,2,24] In addition, it satisfies a property of monotonicity, which is not satisfied by (or even applicable to) the Shannon entropy, but which is one of the properties of the Hartley uncertainty.

A measure of conditional uncertainty for possibility distributions is also derived in this paper. The measure, referred to as the *conditional U-uncertainty* and defined by Eqs. (17) and (18), is related to the U-uncertainties of the joint and marginal possibility distributions in exactly the same way as their counterparts for the Shannon entropy and the corresponding probability distributions.

The conditional U-uncertainty is derived in this paper without any use of the notion of conditional possibilities. The meaning of conditional possibilities is currently a controversial issue[26,32] which we hope the definition of the conditional U-uncertainty presented here may help to resolve.

When the information obtained from the outcome of an experiment is defined in terms of uncertainty associated with the relevant set of potential outcomes prior to the experiment, then any meaningful measure of uncertainty is a measure of information as well. Hence, the U-uncertainty is also a measure of information within the framework of possibility theory; we may therefore refer to it as the *U-information.*

It is clear from the discussions and results in this paper that the U-uncertainty (or U-information) plays basically the same role within possibility theory as does the Shannon entropy within probability theory. As such, it provides us with a base for developing an alternative information theory, one formulated in terms of possibility theory rather than probability theory.

Due to the perfect analogy between the U-uncertainty/information and Shannon entropy, meaningful possibilistic counterparts can easily be derived for the various well known laws and principles of information expressed by the Shannon entropy.[4,19,42] Such laws and principles based on the U-uncertainty should prove as useful for dealing with possibilistic systems as their well established counterparts have proved to be for probabilistic systems.

One such principle, which has been recognized within probability theory as an important and well justified principle of inductive inference, is the principle of maximum entropy.[15,16,27,38] Another principle, which is important for pattern recognition and data classification is the principle of minimum entropy.[16,43] For similar problems formulated in terms of possibility theory, these principles can now be reformulated as principles of maximum U-uncertainty and minimum U-uncertainty, respectively.

One issue associated with possibilistic uncertainty which remains unresolved is the question of the uniqueness of the U-uncertainty. Is it the only function that has the required properties $(V1)$–$(V10)$? At this time, the question can be answered only partially. It is clear from the uniqueness of the Hartley uncertainty[36] that any possibilistic uncertainty must be expressed in terms of the logarithmic functions of the relevant ℓ-cuts. However, it is not clear whether the manner in which the logarithms of the ℓ-cuts are aggregated in the U-uncertainty (Eq. (12)) is the only way to derive a possibilistic uncertainty measure that possesses peoperties $(V1)$–$(V10)$.

ACKNOWLEDGEMENT

The authors would like to acknowledge the contributions to this paper of Abdul Hai and Michael Pittarelli, graduate students in the Ph.D. program of the School of Advanced Technology, SUNY-Binghamton. A number of measures which were considered as candidates for the possibilistic uncertainty, including the Yager measure, were excluded by the results of simulation experiments performed by Abdul Hai. This helped the authors to eventually pinpoint the measure introduced and justified in this paper. Michael Pittarelli, on the other hand, was quite instrumental in reviewing the original version of this paper and suggesting a number of improvements.

REFERENCES

1. J. Aczél, and Z. Daróczy, *On Measures of Information and Their Characterizations.* Academic Press, New York, 1975.
2. J., Aczél, B. Forte and C.T. Ng, "Why the Shannon and Hartley entropies are 'natural'." *Advances in Applied Probability*, 6, 1974, pp. 131–146.
3. W. R. Ashby, "Measuring the internal information exchange in a system." *Cybernetica*, 1, No. 1, 1965, pp. 5–22.
4. W. R. Ashby, "Two tables of identities governing information flows within large systems." *ASC Communications*, 1, No. 2, 1969, pp. 3–8.
5. W. R. Ashby, "Information flows within co-ordinated systems." In: *Progress of Cybernetics*, edited by J. Rose, Gordon and Breach, London, 1970, pp. 57—64.
6. W. R. Ashby, "Systems and their informational measures." In: *Trends in General Systems Theory*, edited by G. J. Klir, John Wiley, New York, 1972, pp. 78–97.
7. D. E. Boekee and J. C. A. Van Der Lubbe, "The R-norm information measure." *Information and Control*, 45, No. 2, 1980, pp. 136–155.
8. L. Brillouin, *Science and Information Theory.* Academic Press, New York, 1956.
9. L. Brillouin, *Scientific Uncertainty and Information.* Academic Press, New York, 1964.

10. G. Broekstra, "On the representation and identification of structure systems." *International Journal of Systems Science*, **9**, No. 11, 1978, pp. 1271–1293.
11. R. E. Cavallo and G. J. Klir, "A conceptual foundation for systems problem solving." *International Journal of Systems Science*, **9**, No. 2, 1978, pp. 219–236.
12. R. E. Cavallo and G. J. Klir, "Systems of overall possibilities: reconstruction aspects." In: *Applied Systems and Cybernetics*, edited by G. E. Lasker, Pergamon Press, New York, 1981, pp. 939–946.
13. R. E. Cavallo and G. J. Klir, "Reconstruction of possibilistic behavior systems." *Fuzzy Sets and Systems*, **8**, No. 2, August 1982, pp. 175–197.
14. C. Cherry, *On Human Communication*. The M.I.T. Press, Cambridge, Mass., 1957.
15. R. Christensen, *Foundations of Inductive Reasoning*. Entropy, Lincoln, Mass., 1980.
16. R. Christensen, *Entropy Minimax Source Sourcebooks*, Vols. 1–4. Entropy, Lincoln, Mass., 1981.
17. R. C. Conant, "Detecting subsystems of a complex system." *IEEE Trans. on Systems, Man, and Cybernetics*, **SMC-2**, No. 4, 1972, pp. 550–553.
18. R. C. Conant, "Measuring the strength of intervariable relations." *Kybernetes*, **2**, No. 1, 1973, pp. 47–52.
19. R. C. Conant, "Laws of information which govern systems." *IEEE Trans. on Systems, Man, and Cybernetics*, **SMC-6**, No. 4, 1976, pp. 240–255.
20. R. C. Conant, "Structural modelling using a simple information measure." *International Journal of Systems Science*, **11**, No. 6, June 1980, pp. 721–730.
21. Z. Daróczy, "Generalized information functions." *Information and Control*, **16**, No. 1, 1970, pp. 36–51.
22. D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications*. Academic Press, New York, 1980.
23. B. Forte, "On the amount of information given by an experiment." In: *Proc. of the Colloquium on Information Theory*, edited by A. Rényi, Janus Bolyai Mathematical Society, Budapest, 1968, pp. 149–166 (Vol. 1).
24. B. Forte, "Why Shannon entropy." *Symposia Mathematica*, XV, Academic Press, New York, 1975, pp. 137–152.
25. R. V. L. Hartley, "Transmission of information." *The Bell Systems Technical Journal*, **7**, 1928, pp. 535–563.
26. E. Hisdale, "Conditional possibilities, independence and noninteraction." *Fuzzy Sets and Systems*, **1**, No. 4, 1978, pp. 283–297.
27. E. T. Jaynes, "Where do we stand on maximum entropy?" In: *The Maximum Entropy Formalism*, edited by R. L. Levine and M. Tribus, MIT Press, Cambridge, Mass., 1979, pp. 15–118.
28. G. J. Klir, On the representation of activity arrays. *International Journal of General Systems*, **2**, No. 3, 1975, pp. 149–168.
29. G. J. Klir, "General systems problem solving methodology." In: *Modelling and Simulation Methodology*, edited by B. Zeigler *et al.*, North-Holland, Amsterdam, 1979.
30. A. N. Kolmogorov, *Foundations of the Theory of Probability*. Chelsea, New York, 1950.
31. A. N. Kolmogorov, "Three approaches to the quantitative definition of information." *Problems of Information Transmission*, **1**, No. 1, 1965, pp. 1–7.
32. H. T. Nguyen, "On conditional possibility distributions." *Fuzzy Sets and Systems*, **1**, No. 4, October 1978, pp. 299–309.
33. H. T. Nguyen, "Toward a calculus of the mathematical notion of possibility." In: *Advances in Fuzzy Set Theory and Applications*, edited by M. M. Gupta, R. K. Ragade, and R. R. Yager, North-Holland, New York, 1979, pp. 235–246.
34. M. L. Puri and D. Ralescu, "A possibility measure is not a fuzzy measure." *Fuzzy Sets and Systems*, **7**, No. 3, 1982, pp. 311–313.
35. A. Rényi, "On measures of entropy and information." *Proc. 4th Berkeley Symp. on Math. Stat. and Prob.*, Univ. of California Press, Berkeley, 1960, pp. 547–561.
36. A., Rényi, *Probability Theory*. North-Holland, Amsterdam, 1970 (Appendix: "Introduction to information theory," pp. 540–616.
37. C. E. Shannon, "A mathematical theory of communication." *The Bell System Technical Journal*, **27**, July and October, 1948, pp. 379–423, 623–656.
38. J. E. Shore and R. W. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy." *IEEE Trans. on Information Theory*, **IT-26**, No. 1, 1980, pp. 26–37.
39. *Special Issue of the International Journal of General Systems on Reconstructability Analysis*, Vol. 7, No. 1, 1981 (Broekstra, Cavallo, Conant, Klir, Krippendorff).
40. M. Sugeno, "Fuzzy measures and fuzzy integrals: a survey." In: *Fuzzy Automata and Decision Processes*, edited by M. M. Gupta, North-Holland, New York, 1977, pp. 89–102.
41. S. Watanabe, "Information theoretical analysis of multivariate correlations." *IBM Journal of Research and Development*, **4**, No. 1, 1960, pp. 66–82.
42. S. Watanabe, *Knowing and Guessing*. John Wiley, New York, 1969.
43. S. Watanabe, "Pattern recognition as a guest for minimum entropy." *Pattern Recognition*, **13**, No. 5, 1981, pp. 381–387.
44. R. R. Yager, "Measuring tranquility and anxiety in decision making: an application of fuzzy sets." *Technical Report No. RRY* 79–03. Iona College, School of Business Adminstration, New Rochelle, New York, 1979.
45. R. R. Yager, "Measuring tranquility and anxiety in decision making." *International Journal of General Systems*, **8**, No. 3, 1982, pp. 139–146.
46. R. R. Yager, "Aspects of possibilistic uncertainty." *Intern. J. of Man–Machine Studies*, **12**, No. 3, April 1980, pp. 283–298.
47. R. R. Yager, "A foundation for a theory of possibility." *Journal of Cybernetics*, **10**, Nos. 1–3, 1980, pp. 177–204.
48. L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility." *Fuzzy Sets and Systems*, **1**, No. 1, 1978, pp. 3–28.
49. L. A. Zadeh, "PRUF—a meaning representation language for natural languages." *International Journal of Man–Machine Studies*, **10**, No. 4, 1978, pp. 395–460.

For biographies and photographs of the authors, please see Volume 8, No. 3, 1982, p. 180 of this journal.