Annals of Mathematics and Artificial Intelligence **32:** 5–33, 2001.

On measuring uncertainty and uncertainty-based information: Recent developments

George J. Klir and Richard M. Smith

Center for Intelligent Systems and Department of Systems Science and Industrial Engineering, Binghamton University – SUNY, Binghamton, NY 13902, USA

It is shown in this paper how the emergence of fuzzy set theory and the theory of monotone measures considerably expanded the framework for formalizing uncertainty and suggested many new types of uncertainty theories. The paper focuses on issues regarding the measurement of the amount of relevant uncertainty (predictive, prescriptive, diagnostic, etc.) in nondeterministic systems formalized in terms of the various uncertainty theories. It is explained how information produced by an action can be measured by the reduction of uncertainty produced by the action. Results regarding measures of uncertainty (and uncertainty-based information) in possibility theory, Dempster–Shafer theory, and the various theories of imprecise probabilities are surveyed. The significance of these results in developing sound methodological principles of uncertainty and uncertainty-based information is discussed.

Keywords: uncertainty, uncertainty-based information, nonspecificity, conflict, possibility theory, Dempster–Shafer theory, theories of imprecise probabilities, fuzzy sets, monotone measures, rough sets

1. Introduction

It is our contention that scientific knowledge is organized, by and large, in terms of *systems* of various kinds. Each of these systems has been constructed for some *purpose*. The most typical purposes for constructing systems are: prediction, retrodiction, extrapolation in space or within a population, prescription, control, decision making, planning, scheduling, and diagnosis.

In general, every system involves a *relation* among some variables. This relation is utilized, in a purposeful way, for determining unknown states of some variables on the basis of known states of other variables. If the unknown states are always determined uniquely, the system is called *deterministic*; otherwise, it is called *nondeterministic*.

Nondeterministic systems are by far more prevalent in modern science than deterministic ones. A special feature of nondeterministic systems is that they inevitably involve some uncertainty. In each nondeterministic system, the uncertainty is associated with the purpose for which the system was constructed. We distinguished thus predictive uncertainty, retrodictive uncertainty, diagnostic uncertainty, etc. The relevant uncertainty (predictive, prescriptive, etc.) must be properly incorporated into the formal description of each nondeterministic system.

Uncertainty involved in nondeterministic systems is the subject of this paper. To cover this subject comprehensively, we introduce a broad conceptual framework within which various uncertainty theories can be conceived, and we briefly describe how far each of them has been developed thus far.

To develop a fully operational theory for dealing with uncertainty of some conceived type requires that a host of issues be addressed at four distinct levels:

- LEVEL 1 – we need to find an appropriate *mathematical representation* of the conceived type of uncertainty.
- LEVEL 2 – we need to develop a *calculus* by which this type of uncertainty can be properly manipulated.
- LEVEL 3 – we need to find a meaningful way of *measuring* relevant uncertainty in any situation formalizable in the theory.
- LEVEL 4 – we need to develop methodological aspects of the theory, including procedures for making the various *uncertainty principles* operational within the theory.

Our focus in this paper is on level 3. We discuss issues that must be addressed for each uncertainty theory at this level, and present a summary of results as well as open questions at this level. Further details regarding developments in this area prior to 1998 are covered in [24]. More recent developments are examined in [40].

2. Historical background

Traditionally, science, mathematics, and engineering showed virtually no interest in studying uncertainty. In fact, eliminating uncertainty was viewed as one manifestation of progress. This negative attitude towards uncertainty, prevalent prior to the 20th century, was seriously challenged by some developments in the 20th century. These developments include the emergence of statistical mechanics, Heisenberg's uncertainty principle in quantum physics, and Gödel's theorems regarding consistency and completeness of formal systems. In spite of these developments, the traditional attitude towards uncertainty changed too little and too slow during the first half of the century. While uncertainty became recognized as useful (e.g., in statistical mechanics), for a long time it was tacitly assumed that probability theory is the only way to deal with uncertainty.

The assumption that probability theory is capable of capturing any kind of uncertainty was challenged only in the second half of the 20th century by the emergence of two important generalizations in mathematics. One of them is the generalization of classical set theory to *fuzzy set theory*, introduced by Lotfi Zadeh [49]. The second one is the generalization of classical measure theory [13] to the *theory of monotone measures*, first suggested by Gustave Choquet [3] in his *theory of capacities*. These generalizations enlarged substantially the framework for formalizing uncertainty and allowed us to formalize numerous new types of uncertainty. To explain this enlarged framework, we need to explain the two underlying concepts first: fuzzy sets and monotone measures.

3. Fuzzy sets and monotone measures

3.1. Fuzzy sets

Fuzzy sets are defined on any given universal set of concern by functions analogous to characteristic functions of classical sets. These functions are called *membership functions*. Each membership function defines a fuzzy set on a given universal set by assigning to each element of the universal set its membership grade in the fuzzy set. The set of all membership grades must be at least partially ordered, but it is usually required to form a complete lattice. The most common fuzzy sets, usually referred to as *standard fuzzy sets*, are defined by membership grades in the unit interval [0, 1]. Those for which the maximum (or supremum) is 1 are called *normal*. Fuzzy sets that are not normal are called *subnormal*.

Two distinct notations are most commonly employed in the literature to denote membership functions. In one of them, the membership function of a fuzzy set A is denoted by μ_A and, assuming that A is a standard fuzzy set, its form is

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

where X denotes the universal set of concern. In the second notation, the membership function is denoted by A and has, of course, the same form

$$A : X \to [0, 1].$$

According to the first notation, the symbol of the fuzzy set involved is distinguished from the symbol of its membership function. According to the second notation, this distinction is not made, but no ambiguity results from this double use of the same symbol since each fuzzy set is completely and uniquely defined by one particular membership function. The second notation is adopted in this paper; it is simpler and, by and large, more popular in the current literature on fuzzy sets.

Contrary to the symbolic role of numbers 0 and 1 in characteristic functions of classical sets, numbers assigned to relevant objects by membership functions of fuzzy sets have a numerical significance. This significance is preserved when classical sets are viewed (from the standpoint of fuzzy set theory) as special fuzzy sets, often referred to as *crisp sets*.

An important property of fuzzy sets is their capability to express gradual transitions from membership to nonmembership. This expressive capability has great utility. For example, it allows us to capture, at least in a crude way, the meaning of expressions in natural language, most of which are inherently vague. Crisp sets are hopelessly inadequate for this purpose. However, it is important to realize that meanings of expressions in natural language are strongly dependent on the context within which they are used.

Among the most important concepts associated with standard fuzzy sets is the concept of an *α-cut*. Given a fuzzy set A defined on X and a number α in the unit interval

[0, 1], the α-cut of A, denoted by ${}^{\alpha}A$, is the crisp set that consists of all elements of X whose membership degrees in A are greater than or equal to α; that is,

$${}^{\alpha}A = \{x \in X \mid A(x) \geqslant \alpha\}.$$

It is obvious that α-cuts are classical (crisp) sets, which for any given fuzzy set A form a nested family of sets in the sense that

$${}^{\alpha}A \subseteq {}^{\beta}A$$

when $\alpha > \beta$.

Each fuzzy set is uniquely represented by the family of its α-cuts [25]. Any property that is generalized from classical set theory into the domain of fuzzy set theory by requiring that it holds in all α-cuts in the classical sense is called a *cutworthy property*.

Fuzzy set theory is now fairly well developed [25]. Many researchers have contributed to it, but perhaps the most important role in its development, not only in its founding, was played by Zadeh. Fortunately, this role is now well documented by two volumes of his selected papers in the period 1965–1995 [27,47].

3.2. Monotone measures

Given a universal set X and a non-empty family C of subsets of X (usually with an appropriate algebraic structure), a *monotone measure*, g, on $\langle X, \mathsf{C}\rangle$ is a function

$$g : \mathsf{C} \to [0, \infty]$$

that satisfies the following requirements:

(g1) $g(\varnothing) = 0$ (vanishing at the empty set);

(g2) for all $A, B \in \mathsf{C}$, if $A \subseteq B$, then $g(A) \leqslant g(B)$ (monotonicity);

(g3) for any increasing sequence $A_1 \subseteq A_2 \subseteq \cdots$ of sets in C,

$$\text{if } \bigcup_{i=1}^{\infty} A_i \in \mathsf{C}, \quad \text{then} \quad \lim_{i\to\infty} g(A_i) = g\left(\bigcup_{i=1}^{\infty} A_i\right)$$

(continuity from below);

(g4) for any decreasing sequence $A_1 \supseteq A_2 \supseteq \cdots$ of sets in C,

$$\text{if } \bigcap_{i=1}^{\infty} A_i \in \mathsf{C}, \quad \text{then} \quad \lim_{i\to\infty} g(A_i) = g\left(\bigcap_{i=1}^{\infty} A_i\right)$$

(continuity from above).

Functions that satisfy requirements (g1), (g2), and either (g3) or (g4) are equally important in the theory of monotone measures. These functions are called *semicontinuous* from below or above, respectively. When the universal set X is finite, requirements (g3) and (g4) are trivially satisfied and may thus be disregarded. If $X \in \mathsf{C}$ and $g(X) = 1$, g is called a *regular* monotone measure (or *regular* semicontinuous monotone measure).

For any pair $A, B \in \mathsf{C}$ such that $A \cap B = \varnothing$, a monotone measure g is capable of capturing any of the following situations:

(a) $g(A \cup B) > g(A) + g(B)$, which expresses a cooperative action or synergy between A and B in terms of the measured property;

(b) $g(A \cup B) = g(A) + g(B)$, which expresses the fact that A and B are noninteractive with respect to the measured property;

(c) $g(A \cup B) < g(A) + g(B)$, which expresses some sort of inhibitory effect or incompatibility between A and B as far as the measured property is concerned.

Observe that probability theory, which is based on classical measure theory [13] is capable of capturing only situation (b). This demonstrates that the theory of monotone measures provides us with a considerably broader framework than probability theory for formalizing uncertainty. As a consequence, it allows us to capture types of uncertainty that are beyond the scope of probability theory.

For some historical reasons of little significance, monotone measures are often referred to in the literature as *fuzzy measures* [43]. This name is somewhat confusing since no fuzzy sets are involved in the definition of monotone measures. To avoid this confusion, the term "fuzzy measures" should be reserved to measures (additive or non-additive) that are defined on families of fuzzy sets.

4. Framework for formalizing uncertainty

The emergence of fuzzy set theory and the theory of monotone measures considerably expanded the framework for formalizing uncertainty. This expansion is two-dimensional. In one dimension, the formalized language of classical set theory is expanded to the more expressive language of fuzzy set theory, where further distinctions are based on special types of fuzzy sets. In the other dimension, the classical (additive) measure theory is expanded to the less restrictive theory of monotone measures, within which various branches can be distinguished by monotone measures with different special properties.

The two-dimensional expansion of possible uncertainty theories is illustrated by the matrix in figure 1. In addition to classical sets and fuzzy sets, also listed under the formalized languages in this figure are rough sets and their combinations with fuzzy sets. *Rough sets*, which were introduced by Pawlak [32], are basically imprecise representations of classical sets in terms of a partition (induced by an equivalence relation) defined on the universal set involved. Each rough set represents a given set by two subsets of the partition: a *lower approximation*, which consists of all blocks of the partition that are included in the represented set, and an *upper approximation*, which consists of all blocks that overlap with the represented set.

UNCERTAINTY THEORIES			FORMALIZED LANGUAGES				
			CLASSICAL SETS	NONCLASSICAL SETS			
				FUZZY SETS	ROUGH SETS	FUZZY ROUGH SETS	ROUGH FUZZY SETS
MONOTONE MEASURES	ADDITIVE	CLASSICAL NUMERICAL PROBABILITY	1	2			
	NONADDITIVE	POSSIBILITY/ NECESSITY	3	4			
		BELIEF/ PLAUSIBILITY	5	6			
		CAPACITIES OF VARIOUS ORDERS	7				
		COHERENT LOWER AND UPPER PROBABILITIES	8				
		CLOSED CONVEX SETS OF PROBABILITY DISTRIBUTIONS	9				
		. . .					

Figure 1. Classification of uncertainty theories.

Fuzzy sets and rough sets model different types of uncertainty. Since both types are relevant in some applications, it is useful to combine them as argued by Dubois and Prade [10]. Rough sets based on a fuzzy equivalence relation [25] are usually called *fuzzy rough sets*, while rough-set approximations of α-cuts of fuzzy sets in terms of a given crisp equivalence relation are called *rough fuzzy sets*. Again, these two combinations model different aspects of uncertainty and, consequently, have different domains of applicability.

Under the entry of non-additive measures in figure 1, only a few representative types are listed. Some of them are presented as pairs of dual measures employed jointly in various uncertainty theories, such as possibility and necessity measures in possibility theory [7,9] or belief and plausibility measures in the Dempster–Shafer theory of evidence [38] (see also section 5).

An uncertainty theory of a particular type is formed by choosing a formalized language of a particular type and expressing relevant uncertainty (predictive, prescriptive, diagnostic, etc.) involved in situations described in this language in terms of a measure (or a pair of measures) of a certain type. This means that each entry in the matrix in figure 1 represents an uncertainty theory of a particular type. For example, the labelled entries in figure 1 represent the following theories:

1. Classical probability theory [28].

2. Probability theory based on fuzzy events [50].

3. Classical (crisp) possibility theory (section 5.1) [33,45].

4. Fuzzy-set interpretation of possibility theory [9,21,51].

5. Dempster–Shafer theory of evidence [38].

6. Fuzzified Dempster–Shafer theory [48].

7–9. Various formalizations of imprecise probabilities [41,42].

Thus far, important results regarding measures of uncertainty (at level 3) have been obtained in possibility theory, Dempster–Shafer theory, and the theories of imprecise probabilities identified in figure 1 by labels 7–9 (in addition to classical results in probability theory). To discuss these results, we need to introduce basic notions of these uncertainty theories first.

5. Theories of uncertainty

5.1. *Possibility theory*

It was first recognized by Zadeh [51] that possibility theory is a natural tool for representing and manipulating information expressed in terms of fuzzy propositions. In this interpretation of possibility theory, the classical (crisp) possibility and necessity measures based upon modal logic are extended to their fuzzy counterparts via the α-cut representation [33,45].

Consider a set of alternatives, X. One of the alternatives is true, but we are not certain which one it is, due to limited evidence. Assume that we only know, according to all evidence available, that it is not possible that the true alternative be outside a given set E, where $\varnothing \neq E \subseteq X$. This simple evidence can be expressed by a classical (crisp) *possibility measure*, Pos_E, defined on X by the formulas

$$\mathrm{Pos}_E(\{x\}) = \begin{cases} 1, & \text{when } x \in E, \\ 0, & \text{when } x \in \overline{E}, \end{cases}$$

for all $x \in X$ and

$$\mathrm{Pos}_E(A) = \sup_{x \in A} \mathrm{Pos}_E(\{x\}) \tag{1}$$

for all $A \in \mathsf{P}(X)$. Associated with the possibility measure Pos_E is a *necessity measure*, Nec_E, defined via the equation

$$\mathrm{Nec}_E(A) = 1 - \mathrm{Pos}_E(\overline{A})$$

for all $A \in \mathsf{P}(X)$, where $\overline{A}$ denotes the complement of A.

Assume now that the set E, in terms of which the evidence is expressed, is a standard and normal fuzzy set. Then, the previous formulas are still applicable to the α-cuts of E, provided that $\varnothing \neq {}^{\alpha}E \subseteq X$ for all $\alpha \in [0, 1]$. For each $\alpha \in [0, 1]$, we can define a possibility measure ${}^{\alpha}\mathrm{Pos}_E$ in the same way as before:

$$ {}^{\alpha}\mathrm{Pos}_E(\{x\}) = \begin{cases} 1, & \text{when } x \in {}^{\alpha}E, \\ 0, & \text{when } x \in \overline{{}^{\alpha}E}, \end{cases} \tag{2} $$

for all $x \in X$ and

$$ {}^{\alpha}\mathrm{Pos}_E(A) = \sup_{x \in A} {}^{\alpha}\mathrm{Pos}_E(\{x\}) $$

for all $A \in \mathsf{P}(X)$. Now using the α-cut representation of E [25], we have

$$ E(x) = \sup_{\alpha \in [0,1]} \min[\alpha, {}^{\alpha}E(x)] $$

for all $x \in X$, where ${}^{\alpha}E$ denotes here the characteristic function of the α-cut of E. Since equation (2) can be rewritten as

$$ {}^{\alpha}\mathrm{Pos}_E(\{x\}) = {}^{\alpha}E(x) $$

for all $\alpha \in [0, 1]$ and $x \in X$, it is natural to define a possibility measure, Pos_E, in terms of the possibility measures ${}^{\alpha}\mathrm{Pos}_E$ ($\alpha \in [0, 1]$) via the α-cut representation of E. Hence,

$$ \mathrm{Pos}_E(\{x\}) = \sup_{\alpha \in [0,1]} \min[\alpha, {}^{\alpha}\mathrm{Pos}_E(\{x\})] $$

for all $x \in X$, which can be rewritten as

$$ \mathrm{Pos}_E(\{x\}) = E(x). \tag{3} $$

This definition of possibility measures, which is due to Zadeh [51], is usually referred to in the literature as the *standard fuzzy-set interpretation of possibility theory*. Given $\mathrm{Pos}_E(\{x\})$ for all $x \in X$, $\mathrm{Pos}_E(A)$ is then calculated by the equation

$$ \mathrm{Pos}_E(A) = \sup_{x \in A} \mathrm{Pos}_E(\{x\}) \tag{4} $$

for all $A \in \mathsf{P}(X)$. When A is a fuzzy set, equation (4) must be replaced with the more general equation

$$ \mathrm{Pos}_E(A) = \sup_{x \in X} \min[A(x), \mathrm{Pos}_E(\{x\})]. \tag{5} $$

The associated necessity measure, Nec_E, is then defined by the equation

$$ \mathrm{Nec}_E(A) = 1 - \mathrm{Pos}_E(\overline{A}) \tag{6} $$

for each set A, which may be crisp or fuzzy; if A is fuzzy, then $\overline{A}(x) = 1 - A(x)$ for all $x \in X$.

Possibility measures and the associated necessity measures that represent evidence expressed in terms of standard fuzzy set via equations (3)–(6) are thus cutworthy. They form a coherent theory of evidence that is referred to as *possibility theory* [7,9]. The requirement that

$$\varnothing \neq {}^{\alpha}E \subseteq X$$

for all $\alpha \in [0, 1]$ means that E must be a normal fuzzy set.

When evidence is expressed in terms of a subnormal fuzzy set, the coherence of the standard fuzzy-set interpretation of possibility theory (the one just described) is lost. Observe, for example, that

$$\mathrm{Pos}_E\bigl\{x \in X \mid E(x) > 0\bigr\} = h_E,$$

where

$$h_E = \sup_{x \in X} E(x),$$

and

$$\mathrm{Nec}_E\bigl\{x \in X \mid E(x) > 0\bigr\} = 1.$$

When $h_E < 1$, then

$$\mathrm{Nec}_E\bigl\{x \in X \mid E(x) > 0\bigr\} > \mathrm{Pos}_E\bigl\{x \in X \mid E(x) > 0\bigr\}.$$

This violates the fundamental inequality of possibility theory, $\mathrm{Nec}_E(A) \leqslant \mathrm{Pos}_E(A)$, which is required to hold for all $A \in \mathsf{P}(X)$.

The fact that the standard fuzzy-set interpretation of possibility theory, defined by equation (3), is not applicable to subnormal fuzzy sets has been recognized in the literature since the mid 1980s. In a recent paper [21], it is shown that the only way to make a fuzzy-set interpretation applicable to all standard fuzzy sets without distorting given evidence is to replace equation (3) with the more general equation

$$\mathrm{Pos}_E\bigl(\{x\}\bigr) = E(x) + 1 - h_E \tag{7}$$

for $x \in X$.

At this point, it is appropriate to make a remark regarding the introduced general fuzzy-set interpretation of possibility theory. In this interpretation, possibility theory is based on two dual semicontinuous monotone measures: possibility measures (semicontinuous from below) and necessity measures (semicontinuous from above). The theory may be characterized in terms of either of these measures; the other one is then defined via equation (6). For example, given a universal set X, a possibility measure, Pos, is a function

$$\mathrm{Pos} : \mathsf{P}(X) \to [0, 1]$$

that is characterized by the following properties:

(a) $\mathrm{Pos}(\varnothing) = 0$;

(b) $\mathrm{Pos}(X) = 1$;

(c) for any family $\{A_i \mid A_i \in \mathsf{P}(X),\ i \in I\}$, where I is an arbitrary index set,

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

In a more general formulation [7], which is not followed in this paper, Pos may be viewed as a function from an ample field on X (a family of subsets of X that is closed under arbitrary unions and intersections, and under complementation in X) to a given complete lattice.

An alternative way of using a generalized version of possibility measures, one for which property (b) is not required, for formalizing reasoning under incomplete and/or partially inconsistent knowledge has recently been pursued under the name *possibilistic logic* [11]. In this area, which is not considered in this paper, the difference $\mathrm{Nec}(A) - \mathrm{Pos}(A) > 0$ is viewed as the amount of internal conflict pertaining to A for a subnormal possibility distribution.

5.2. *Dempster–Shafer theory*

Dempster–Shafer theory of evidence (DST) employs a non-additive measure that is called a belief measure. Given a universal set X (usually referred to as the frame of discernment in DST), a *belief measure*, Bel, is a function

$$\mathrm{Bel} : \mathsf{P}(X) \to [0, 1]$$

such that $\mathrm{Bel}(\varnothing) = 0$, $\mathrm{Bel}(X) = 1$ and

$$\begin{aligned}\mathrm{Bel}(A_1 \cup A_2 \cup \cdots \cup A_n) \geqslant{}& \sum_j \mathrm{Bel}(A_j) - \sum_{j<k} \mathrm{Bel}(A_j \cap A_k) \\ &+ \cdots + (-1)^{n+1}\,\mathrm{Bel}(A_1 \cap A_2 \cap \cdots \cap A_n) \qquad (8)\end{aligned}$$

for all possible families of subsets of X. Due to this inequality, belief measures are also called monotone capacities of order ∞ [3]. When X is infinite, function Bel is also required to be *continuous from above*.

For any given belief measure Bel, a dual measure, Pl, is defined by the equation

$$\mathrm{Pl}(A) = 1 - \mathrm{Bel}\left(\overline{A}\right)$$

for all $A \in \mathsf{P}(X)$. This measure, which is called a *plausibility measure* (or alternate capacity of order ∞ [3]) satisfies the inequality

$$\begin{aligned}\mathrm{Pl}(A_1 \cap A_2 \cap \cdots \cap A_n) \leqslant{}& \sum_j \mathrm{Pl}(A_j) - \sum_{j<k} \mathrm{Pl}(A_j \cup A_k) \\ &+ \cdots + (-1)^{n+1}\,\mathrm{Pl}(A_1 \cup A_2 \cup \cdots \cup A_n) \qquad (9)\end{aligned}$$

for all possible families of subsets X. Moreover, $\mathrm{Pl}(\varnothing) = 0$ and $\mathrm{Pl}(X) = 1$. When X is infinite, function Pl is also required to be *continuous from below*.

The belief measure Bel and its dual plausibility measure Pl are conveniently characterized by a function

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

which is required to satisfy two conditions:

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

(b) $\sum_{A \in \mathsf{P}(x)} m(A) = 1$.

This function is called in the DST literature a *basic probability assignment*. For each $A \in \mathsf{P}(X)$, the value $m(A)$ expresses the degree of support of the evidential claim that the true alternative (prediction, diagnosis, etc.) is in the set A but not in any special subset of A. Any additional evidence supporting the claim that the true alternative is in a subset of A, say $B \subset A$, must be expressed by another nonzero value $m(B)$. Condition (b) is called a normalization condition of DST. When m is derived from conflicting evidence, the values of $m(A)$ may be such that

$$\sum_{A \in \mathsf{P}(x)} m(A) = h,$$

where $h < 1$. As shown by Yager [46], it is meaningful to normalize m in these cases by replacing $m(X)$ with $m(X) + 1 - h$.

Given a particular basic probability assignment m, the corresponding belief measure and plausibility measure are determined for all sets $A \in \mathsf{P}(X)$ by the formulas

$$\mathrm{Bel}(A) = \sum_{B \mid B \subseteq A} m(B), \tag{10}$$

$$\mathrm{Pl}(A) = \sum_{B \mid A \cap B \neq \emptyset} m(B). \tag{11}$$

The three functions, Bel, Pl, and m, are alternative representations of the same evidence. Given any one of them, the other two are uniquely determined. For example,

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} \, \mathrm{Bel}(B) \tag{12}$$

for all $A \in \mathsf{P}(X)$ [38]. For any given belief measure Bel, the unique function m obtained by equation (12) is called a *Möbius inverse* of Bel.

Given a basic probability assignment m, every set for which $m(A) > 0$ is called a *focal element*. The pair $\langle \mathsf{F}, m \rangle$, where F denotes the set of all focal elements induced by m is called a *body of evidence*.

Given two frames of discernment, X and Y, a basic probability assignment of the form

$$m : \mathsf{P}(X \times Y) \to [0, 1],$$

which expresses evidence pertaining to the frame of discernment $X \times Y$, is called a *joint basic probability assignment*. Each focal element induced by m is a binary relation R on $X \times Y$. When R is projected on set X and set Y, we obtain the sets

$$R_X = \{x \in X \mid \langle x, y\rangle \in R \text{ for some } y \in Y\},$$
$$R_Y = \{y \in Y \mid \langle x, y\rangle \in R \text{ for some } x \in X\},$$

respectively. These sets are then used for defining *marginal basic probability assignments*, m_X and m_Y, of m via formulas

$$m_X(A) = \sum_{R \mid R_X = A} m(R) \quad \text{for all } A \in \mathsf{P}(X),$$
$$m_Y(B) = \sum_{R \mid R_Y = B} m(R) \quad \text{for all } B \in \mathsf{P}(Y).$$

Bodies of evidence $\langle \mathsf{F}_X, m_X\rangle$ and $\langle \mathsf{F}_Y, m_Y\rangle$ are, according to DST, noninteractive if and only if for all $A \in \mathsf{F}_X$ and all $B \in \mathsf{F}_Y$

$$m(A \times B) = m_X(A) m_Y(B)$$

and $m(R) = 0$ for all $R \neq A \times B$.

When all focal elements in a given body of evidence are singletons, the associated belief measure and plausibility measure collapse into a single measure that is formally equivalent to the classical probability measure, which is additive. DST is best covered in the book by Shafer [38].

5.3. *Theories of imprecise probabilities*

Several theories of imprecise probabilities have been developed over the last several decades. These theories can be partially ordered by their levels of generality. Among the well-developed theories of imprecise probabilities, the *theory based on coherent lower and upper previsions* (developed by Walley [41]) is currently the most general one. This theory contains the less general *theory based on coherent lower and upper probabilities*. These are, respectively, a superadditive measure $\underline{g}$ that is continuous form above and a subadditive measure $\overline{g}$ that is continuous from below. That is, $\underline{g}$ and $\overline{g}$ are monotone measures for which

$$\begin{aligned} \underline{g}(A \cup B) &\geqslant \underline{g}(A) + \underline{g}(B), \\ \overline{g}(A \cup B) &\leqslant \overline{g}(A) + \overline{g}(B), \end{aligned} \tag{13}$$

for all $A, B \in \mathsf{P}(X)$, assuming (as in DST and possibility theory) that $\underline{g}$ and $\overline{g}$ are defined on $\mathsf{P}(X)$ and X is finite.

The theory based on coherent lower and upper probabilities contains, in turn, the less general theory based on *capacities of order* 2 [3]. Here, the lower and upper probabilities, $\underline{g}$ and $\overline{g}$, are monotone measures for which

$$\begin{aligned}\underline{g}(A \cup B) &\geqslant \underline{g}(A) + \underline{g}(B) - \underline{g}(A \cap B),\\ \overline{g}(A \cap B) &\leqslant \overline{g}(A) + \overline{g}(B) - \overline{g}(A \cup B),\end{aligned} \tag{14}$$

for all $A, B \in \mathsf{P}(X)$. This theory is equally general as another uncertainty theory, one that is based on *feasible interval-valued probability distributions* [31,44].

Less general uncertainty theories are then based on *capacities of order k* [3]. For each $k > 2$, the lower and upper probabilities, $\underline{g}$ and $\overline{g}$, satisfy the inequalities

$$\begin{aligned}\underline{g}\left(\bigcup_{j=1}^{k} A_j\right) &\geqslant \sum_{\substack{K \subseteq N_k,\\ K \neq \emptyset}} (-1)^{|K|+1} \underline{g}\left(\bigcap_{j \in K} A_j\right),\\ \overline{g}\left(\bigcap_{j=1}^{k} A_j\right) &\leqslant \sum_{\substack{K \subseteq N_k,\\ K \neq \emptyset}} (-1)^{|K|+1} \overline{g}\left(\bigcup_{j \in K} A_j\right)\end{aligned} \tag{15}$$

for all families of k subsets of X, where $N_k = \{1, 2, \ldots, k\}$. Clearly, if $k' > k$ then the theory based on capacities of order k' is less general than the one based on capacities of order k. The least general of all these theories is the one in which the inequalities (15) are required to hold for all $k \geqslant 2$ (the underlying capacity is said to be of order ∞). This theory is, of course, DST, which is introduced in section 5.2.

One additional theory of imprecise probabilities should be mentioned, which is less general than DST. It is the theory based on Sugeno λ-measures, g_λ, which are characterized by the equation

$$g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A) g_\lambda(B), \tag{16}$$

for all disjoint sets $A, B \in \mathsf{P}(X)$, where $\lambda \in (-1, \infty)$. As is well known [43], g_λ is a belief measure when $\lambda > 0$, probability measure when $\lambda = 0$, and plausibility measure when $\lambda < 0$. Moreover, given a belief measure g_λ (i.e., $\lambda > 0$), the dual plausibility measure is obtained for $\lambda' = -\lambda/(\lambda + 1)$.

All the mentioned theories of imprecise probabilities are thus linearly ordered by levels of their generality, as shown in figure 2. The most general theory in this sequence is the theory based on lower and upper previsions developed by Walley [41], the least general is the theory based on the Sugeno λ-measures. While classical probability theory is even less general, it does not qualify as a theory of imprecise probabilities.

As shown by Walley [41,42], there is a one-to-one correspondence between coherent lower previsions and nonempty closed convex sets of probability distributions. Hence, the theory based on *closed convex sets of probability distributions*, which has been pursued for some time by Kyburg [29], is equally general as the theory developed by Walley [41], and all the other theories of imprecise probabilities introduced in this

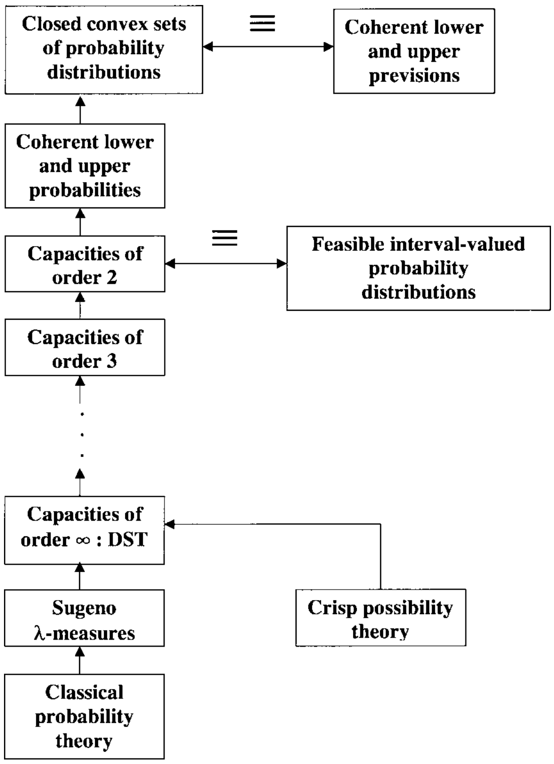

Figure 2. Ordering of theories of imprecise probabilities by levels of their generality.

section are subsumed under it. This is significant since we are now capable of measuring uncertainty associated with closed convex sets of probability distributions (section 6).

Some arguments have already been made in the literature that the theories developed by Walley [41] and Kyburg [29] should be further generalized. Thus, for example, Kyburg and Pittarelli [30] argue for using sets of probability distributions that are not necessarily convex, and Walley [42] argues for two generalizations of his theory. These various prospective generalizations of existing theories are not well developed as yet and, therefore, we do not cover them in this paper.

6. Measures of uncertainty and uncertainty-based information

6.1. Basic issues

In each uncertainty theory, uncertainty is represented by a function that assigns to each set of relevant alternatives (predictions, prescriptions, etc.) a number in the unit interval [0, 1], which expresses the *degree of evidence* (likelihood, belief, plausibility, etc.) that the true alternative is in the set. Let this function be called an *uncertainty function*. Examples of uncertainty functions are probability measures, possibility measures, or belief measures. Uncertainty functions of each uncertainty theory satisfy certain special requirements, which distinguish them from uncertainty functions of other uncertainty theories.

A *measure of uncertainty* of some conceived type in a given uncertainty theory is a *functional* that assigns to each uncertainty function in the theory a nonnegative real number. This number is supposed to measure, in an intuitively meaningful way, the amount of uncertainty of the considered type that is embedded in the uncertainty function. To be acceptable as a measure of the amount of uncertainty of a given type in a particular uncertainty theory, the functional must satisfy several intuitively essential axiomatic requirements. Specific mathematical formulation of each of the requirements depends on the uncertainty theory involved. However, the requirements can be described informally in a generic form, independent of the various uncertainty calculi.

The following axiomatic requirements, each expressed in a generic form, are applicable to all uncertainty theories:

1. *Subadditivity* – the amount of uncertainty in a joint representation of evidence (defined on a Cartesian product) cannot be greater than the sum of the amounts of uncertainty in the associated marginal representations of evidence.
2. *Additivity* – the two amounts of uncertainty considered under subadditivity become equal if and only if the marginal representations of evidence are noninteractive according to the rules of the uncertainty calculus involved.
3. *Range* – the range of uncertainty is $[0, M]$, where 0 must be assigned to the unique uncertainty function that describes full certainty and M depends on the cardinality of the universal set involved and on the chosen unit of measurement.
4. *Continuity* – any measure of uncertainty must be a continuous functional.
5. *Expansibility* – expanding the universal set by alternatives that are not supported by evidence must not affect the amount of uncertainty.
6. *Branching/Consistency* – when uncertainty can be computed in more ways, all intuitively acceptable, the results must be the same (consistent).

The remaining two requirements are applicable only to some theories of uncertainty:

7. *Monotonocity* – when evidence can be ordered in the uncertainty theory employed (as in possibility theory), the relevant uncertainty measure must preserve this ordering.
8. *Coordinate invariance* – when evidence is described within the n-dimensional Euclidean space ($n \geqslant 1$), the relevant uncertainty measure must not change under isometric transformation of coordinates.

When distinct types of uncertainty coexist in a given uncertainty theory, it is not necessary that these requirements be satisfied by each uncertainty type. However, they must be satisfied by an overall uncertainty measure, which appropriately aggregates measures of the individual uncertainty types.

The strongest justification of a functional as a meaningful measure of the amount of uncertainty of a considered type in a given uncertainty theory is obtained when we can prove that it is the only functional that satisfies the relevant requirements and measures the amount of uncertainty in some specific measurement units. A suitable measurement unit is uniquely defined by specifying what the amount of uncertainty should be for a particular (and usually very simple) uncertainty function.

6.2. Uncertainty-based information

In general, uncertainty is an expression of some information deficiency. This suggests that information could be measured in terms of uncertainty reduction. To reduce relevant uncertainty (predictive, retrodictive, prescriptive, diagnostic, etc.) in a situation formalized within a mathematical theory requires that some relevant action be taken by a cognitive agent, such as performing a relevant experiment, searching for a relevant fact, or accepting and interpreting a relevant message. If results of the action taken (an experimental outcome, a discovered fact, etc.) reduce uncertainty involved in the situation, then the amount of information obtained by the action is measured by the amount of uncertainty reduced – the difference between *a priori* and *a posteriori* uncertainty. Measuring information in this way is clearly contingent upon our capability to measure uncertainty within the various mathematical frameworks. Information measured solely by uncertainty reduction is an important, even though restricted, notion of information. To distinguish it from the various other conceptions of information, it is common to refer to it as *uncertainty-based information* [24].

Uncertainty-based information does not capture the rich notion of information in human communication and cognition, but it is very useful in dealing with systems. Given a particular system, it is useful, for example, to measure the amount of information contained in the answer given by the system to a relevant question (concerning various predictions, retrodictions, etc.). This can be done by taking the difference between the amount of uncertainty in the requested answer obtained within the experimental frame of the system (a set of variables and their state sets) in the face of total ignorance and the amount of uncertainty in the answer obtained by the system. This can be written concisely as

$$\text{Information}(A_S \mid S, Q) = \text{Uncertainty}(A_{\mathrm{EF}_S} \mid \mathrm{EF}_S, Q) - \text{Uncertainty}(A_S \mid S, Q),$$

where:

- S denotes a given system,
- EF_S denotes the experimental frame of system S,
- Q denotes a given question,
- A_{EF_S} denotes the answer to question Q obtained solely within the experimental frame EF_S,
- A_S denotes the answer to question Q obtained by system S.

6.3. Nonspecificity

Measurement of uncertainty (and associated information) was first conceived by Hartley [14] in terms of the classical possibility measures introduced in section 5.1 and characterized by equation (1). This kind of uncertainty emerges whenever we have evidence that the true alternative is included in a given finite set E of *possible* alternatives. Hartley showed that the only meaningful way to measure uncertainty in this case is to use a function of the form

$$c \log_b |E|,$$

where $|E|$ denotes the cardinality of E, and b and c are positive constants. Each choice of values of b and c determines the unit in which the uncertainty is measured. Requiring that

$$c \log_b 2 = 1,$$

which is the most common choice, uncertainty is measured in *bits*. One bit of uncertainty is equivalent to uncertainty regarding the truth or falsity of one elementary proposition. Choosing conveniently $b = 2$ implies that $c = 1$, and we obtain a unique functional, H, defined for any possibility measure, Pos_E, by the formula

$$\mathrm{H}(\mathrm{Pos}_E) = \log_2 |E|. \tag{17}$$

This functional is usually referred to as *Hartley measure of uncertainty*. Its uniqueness was proven on axiomatic grounds by Rényi [37].

Observe that uncertainty associated with sets of possible alternatives results from the lack of specificity. Large sets result in less specific predictions, diagnoses, etc., than their smaller counterparts. Full specificity is obtained when only one alternative is possible. This type of uncertainty is thus well characterized by the term *nonspecificity.*

Assuming that X denotes the set of all considered alternatives, clearly,

$$0 \leqslant \mathrm{H}(\mathrm{Pos}_E) \leqslant \log_2 |X|$$

for any $E \in \mathsf{P}(X)$. When the set of considered alternatives is a Cartesian product, which is the case when we deal with nondeterministic systems, we distinguish joint, marginal, and conditional Hartley measures. The relationship among these measures, which holds for all other types of uncertainty covered in this paper, is well established and thoroughly covered in [24].

A natural generalization of the Hartley measure of nonspecificity to the fuzzy-set interpretation of possibility theory (introduced in section 5.1) was developed by Higashi and Klir [15] under the name U-*uncertainty.* For any possibility measure, Pos_E, based on evidence expressed in terms of a normal fuzzy set E of possible alternatives, the U-uncertainty, U, is a functional defined by the formula

$$\mathrm{U}(\mathrm{Pos}_E) = \int_0^1 \log_2 \left|{}^{\alpha}E\right| \,\mathrm{d}\alpha. \tag{18}$$

The uniqueness of this functional for measuring nonspecificity in this case was proven on axiomatic grounds by Klir and Mariano [22].

To cover normal as well as subnormal fuzzy sets, equation (18) must be replaced with the equation

$$\mathrm{U}(\mathrm{Pos}_E) = \int_0^{h_E} \log_2\left|{}^{\alpha}E\right| \mathrm{d}\alpha + (1 - h_E)\log_2|X|, \tag{19}$$

where X denotes the universal set of concern. This is a direct consequence of equation (7).

Shortly after the U-uncertainty was discovered, Dubois and Prade [8] showed how to further generalize it to measure nonspecificity in DST. The proposed measure is a functional, N, defined by the formula

$$\mathrm{N}(m) = \sum_{A \in \mathsf{F}} m(A)\log_2|A| \tag{20}$$

where $\langle \mathsf{F}, m\rangle$ is an arbitrary body of evidence in DST. This functional makes good intuitive sense: it is a weighted average of the Hartley measure of focal elements. Its uniqueness was proven by Ramer [34].

In a recent paper, Abellan and Moral [1] generalized the measure of nonspecificity in DST, expressed by equation (20), to closed convex sets of probability distributions. They obtained this generalization in the following way.

Given a closed convex set D of probability distributions p on a universal set X, the lower probability function $\underline{g}_{\mathsf{D}}$ associated with D is defined for each $A \in \mathsf{P}(X)$ by the formula

$$\underline{g}_{\mathsf{D}}(A) = \inf_{p \in \mathsf{D}} \sum_{x \in A} p(x). \tag{21}$$

Next, the Möbius inverse, m_{D}, of $\underline{g}_{\mathsf{D}}$ is calculated for all $A \in \mathsf{P}(X)$ by the usual formula

$$m_{\mathsf{D}}(A) = \sum_{B \subseteq A} (-1)^{|A-B|}\underline{g}_{\mathsf{D}}(B). \tag{22}$$

The measure of nonspecificity, $\mathrm{N}(\mathsf{D})$, associated with D is then defined by the formula

$$\mathrm{N}(\mathsf{D}) = \sum_{A \in \mathsf{P}(X)} m_{\mathsf{D}}(A)\log_2|A|. \tag{23}$$

Observe that equation (23) has the same form as equation (20), but m_{D} in (23) is not a basic probability assignment. As is well known [2],

$$m_{\mathsf{D}}(\emptyset) = 0 \quad \text{and} \quad \sum_{A \in \mathsf{P}(X)} m_{\mathsf{D}}(A) = 1,$$

but the values $m_{\mathsf{D}}(A)$ are negative for some sets A if $\underline{g}_{\mathsf{D}}$ is not a belief measure.

Abellan and Moral [1] showed that the measure of nonspecificity defined by (23) possesses all the essential mathematical properties required for measures of uncertainty. In particular, they proved that the measure has the following properties:

1. It has the *proper range* $[0, \log_2 |X|]$ when measured in bits; 0 is obtained when D contains a single probability distribution; $\log_2|\mathrm{X}|$ is obtained when D contains all probability distributions on X and thus represents total ignorance.

2. It is *subadditive*: $\mathrm{N}(\mathsf{D}) \leqslant \mathrm{N}(\mathsf{D}_X) + \mathrm{N}(\mathsf{D}_Y)$, where

$$\mathsf{D}_X = \left\{ p_x \mid p_x(x) = \sum_{y \in Y} p(x, y) \text{ for some } p \in \mathsf{D} \right\},$$

$$\mathsf{D}_Y = \left\{ p_y \mid p_y(y) = \sum_{x \in X} p(x, y) \text{ for some } p \in D \right\}.$$

3. It is *additive*: $\mathrm{N}(\mathsf{D}) = \mathrm{N}(\mathsf{D}_X) + \mathrm{N}(\mathsf{D}_Y)$ if and only if D_X and D_Y are not interactive, which means that for all $A \in X$ and all $B \in Y$,

$$m_{\mathsf{D}}(A \times B) = m_{\mathsf{D}_X}(A) m_{\mathsf{D}_Y}(B)$$

   and $m_{\mathsf{D}}(R) = 0$ for all $R \neq A \times B$.

4. It is *monotonic*: if D and D' are closed convex sets of probability distributions on X such that $\mathsf{D} \subseteq \mathsf{D}'$, then $\mathrm{N}(\mathsf{D}) \leqslant \mathrm{N}(\mathsf{D}')$.

The nonspecificity measure defined by (23) is thus mathematically sound. It is applicable to all the theories of imprecise probabilities that are subsumed under the theory based on closed convex sets of probability distributions (section 5.3). However, we still need to develop a conceptual common-sense understanding of this measure, in particular the meaning of negative values of $m_{\mathsf{D}}(A)$.

The Hartley measure H and its generalizations U and N are applicable only to finite sets. A *Hartley-like measure*, HL, applicable to possibility measures based on bounded and convex subsets E of possible alternatives in the n-dimensional Euclidean space $\mathbb{R}^n$ was proposed by Klir and Yuan [26]. It is defined by the formula

$$\mathrm{HL}(\mathrm{Pos}_E) = \min_{t \in T} \ln \left[\prod_{i=1}^{n} \left[1 + \mu(E_{i_t})\right] + \mu(E) - \prod_{i=1}^{n} \mu(E_{i_t}) \right], \tag{24}$$

where μ denotes the Lebesgue measure, T denotes the set of all transformations form one orthogonal coordinate system to another, and A_{i_t} denotes the ith one-dimensional projections of E within the coordinate system t. All required properties of this functional except additivity for $n > 3$ were proven in [26]; additivity for any n was proven later by Ramer [35]. In a recent paper [36], Ramer elaborates on his proof and argues that the functional is applicable even to compact sets that are not necessarily convex.

Functional HL can be also applied to measures U and N. The counterpart of equation (19), obtained by replacing the Hartley measure with the Hartley-like measure, is the equation

$$\mathrm{U}(\mathrm{Pos}_E) = \int_0^{h_E} \mathrm{HL}\big({}^{\alpha}E\big)\,\mathrm{d}\alpha + (1 - h_E)\,\mathrm{HL}(X), \tag{25}$$

which is applicable to any possibility measure Pos_E that is derived from a bounded fuzzy set (relation) on $\mathbb{R}^n$ for some $n \geqslant 1$. Similarly, the counterpart of equation (20) is the equation

$$\mathrm{N}(m) = \sum_{A \in \mathsf{F}} m(A)\,\mathrm{HL}(A), \tag{26}$$

provided that F is finite.

Measuring nonspecificity, which is one type of uncertainty, is thus well established in the uncertainty theories of our concern in this paper. However, probability theory is devoid of nonspecificity: if $\langle \mathsf{F}, m\rangle$ is a probabilistic body of evidence in DST (F consists of singletons), then $\mathrm{N}(m) = 0$. This indicates that probability theory is based on another type of uncertainty and that this uncertainty type coexists with nonspecificity in DST. For reasons explained in the next section, this type of uncertainty is usually referred to as *conflict*.

6.4. *Conflict*

A measure of uncertainty in probability theory is well established. It is expressed by a functional S, which in DST assumes the form

$$\mathrm{S}(m) = -\sum_{x \in X} m\big(\{x\}\big) \log_2 m\big(\{x\}\big). \tag{27}$$

Since the measure was proposed by Shannon [39], it is usually referred to as *Shannon measure* (or *Shannon entropy*). The uniqueness of this functional as a measure of probabilistic uncertainty in bits was proven in numerous ways [24].

To get insight into the type of uncertainty measured by S, we can rewrite equation (27) in the form

$$\mathrm{S}(m) = -\sum_{x \in X} m\big(\{x\}\big) \log_2\Bigg[1 - \sum_{y \neq x} m\big(\{y\}\big)\Bigg]. \tag{28}$$

Here, the term

$$\sum_{y \neq x} m\big(\{y\}\big)$$

expresses the total evidential claim pertaining to focal elements that are different from the focal element $\{x\}$. That is, this term expresses the sum of all evidential claims that fully conflict with the one focusing on $\{x\}$. When using

$$-\log_2\left[1-\sum_{y\neq x} m(\{y\})\right],$$

as in equation (28), the expression of conflict is rescaled from $[0, 1]$ to $[0, \infty)$. The logarithmic function is needed in equation (28) to satisfy the requirement of additivity.

It follows from these observations that the Shannon measure is the mean (expected) value of the conflict among evidential claims within a given probabilistic body of evidence. For this reason, the type of uncertainty represented by probability measures is lately referred to as *conflict*.

Since the early 1980s, studious efforts were made by many researchers to determine a general counterpart of the Shannon measure in DST. Although many intuitively promising functionals have been proposed for this purpose, each of them was found upon closer scrutiny to violate some of the essential requirements (section 6.1). In most cases, it was the requirement of subadditivity that was violated. A historical summary of these unsuccessful efforts is covered in [24], and some more recent efforts are examined in [40].

6.5. Aggregate uncertainty

The long, unsuccessful, and often frustrating search for the Shannon-like measure of uncertainty in DST was replaced in the early 1990s with the search for a justifiable aggregate measure, capturing both nonspecificity and conflict. The first attempts were to add the well-justified measure of nonspecificity with one of the proposed candidates for measuring conflict. Unfortunately, all the resulting functionals were again found to violate some of the essential requirements, predominantly subadditivity.

A measure of total uncertainty in DST that possesses all the required mathematical properties was eventually found (independently by several authors in 1993–1994), but not as a composite of measures of uncertainty of the two types. This aggregate uncertainty measure, AU, is defined for each function Bel defined on $\mathsf{P}(X)$ by the formula

$$\mathrm{AU}(\mathrm{Bel}) = \max_{P_{\mathrm{Bel}}}\left[-\sum_{x\in X} p(x)\log_2 p(x)\right], \tag{29}$$

where the maximum is taken over the set P_{Bel} of all probability distributions p on X that are consistent with the given belief measure Bel, which means that they satisfy the constraint

$$\mathrm{Bel}(A) \leqslant \sum_{x\in A} p(x) \quad \text{for all } A \in \mathsf{P}(X),$$

in addition to the usual axiomatic constraints of probability distributions.

Since the aggregate measure AU is defined in terms of the solution to a nonlinear optimization problem, its practical utility was initially questioned. Fortunately, a relatively simple and fully general algorithm for computing the measure was developed and its correctness proven [24].

Observe that AU defined by (29) can be readily generalized to the theory based on closed convex sets of probability distributions. The generalized formula is

$$\mathrm{AU}(\mathsf{D}) = \max_{p \in \mathsf{D}} \left[\sum_{x \in X} -p(x) \log_2 p(x) \right], \tag{30}$$

where D denotes a given closed convex set of probability distributions on X.

Although the functional AU is acceptable on mathematical grounds as an aggregate measure of uncertainty in DST, possibility theory, and the theory based on closed convex sets of probability distributions, it has a severe shortcoming: it is highly insensitive to changes in evidence. To illustrate this undesirable feature, let us examine a very simple example. Let $X = \{x_1, x_2\}$, $m(\{x_1\}) = a$, $m(\{x_2\}) = b$, and $m(X) = 1 - a - b$, where $a + b \leqslant 1$. Then, $\mathrm{AU}(m) = 1$ for all $a \in [0, 0.5]$ and $b \in [0, 0.5]$. Morevover, when $a > 0.5$, AU is independent of b and, similarly, when $b > 0.5$, AU is independent of a.

6.6. Recent considerations and open problems

On the basis of a critical appraisal of all current results (positive as well as negative) regarding measures of uncertainty in DST and possibility theory, Smith [40] proposed three complementary measures of total uncertainty in these theories. To describe them, let $\overline{\mathrm{S}}$ and $\underline{\mathrm{S}}$ denote, respectively, the maximum and minimum Shannon measure within all probability distributions that are consistent with the given body of evidence. Observe that in this convenient notation $\overline{\mathrm{S}} = \mathrm{AU}$.

The first proposed measure of total uncertainty, TU_1, is defined as a linear combination of $\overline{\mathrm{S}}$ and N,

$$\mathrm{TU}_1 = \delta \overline{\mathrm{S}} + (1 - \delta)\mathrm{N}, \tag{31}$$

where $\delta \in (0, 1)$. This measure has two favorable features. First, it satisfies all essential axiomatic requirements (since both $\overline{\mathrm{S}}$ and N satisfy them). Second, it overcomes the insensitivity of measure AU ($= \overline{\mathrm{S}}$). In particular, it distinguishes total ignorance (when $m(X) = 1$) from the uniform probability distribution by values $\log_2 |X|$ and $\delta \log_2 |X|$, respectively. The choice of the value of δ remains an important open problem regarding this measure. It seems that the proper value of δ should be determined in the context of each interpretation of DST. This can be easily done, for example, by specifying the desired value of TU_1 for the uniform probability distribution. However, it is also conceivable that it may be derived on mathematical grounds.

The second proposed measure, TU_2, is defined as the pair consisting of functionals N and $\overline{\mathrm{S}} - \mathrm{N}$:

$$\mathrm{TU}_2 = \langle \mathrm{N}, \overline{\mathrm{S}} - \mathrm{N} \rangle. \tag{32}$$

It is assumed here that the axiomatic requirements are defined in terms of the sum of the two functionals involved, which is always the well-justified aggregate measure $\overline{\mathrm{S}}$ (=AU). In this sense, the measure satisfies trivially all the requirements. Its advantage is that measures of both types of uncertainty that coexist in DST (nonspecificity and conflict) are expressed explicitly. However, an open problem is to justify that the difference $\overline{\mathrm{S}} - \mathrm{N}$ is a genuine generalization of the Shannon measure in DST, and not only a superficial artifact.

The third proposed measure, TU_3, is defined as the pair

$$\mathrm{TU}_3 = \langle \mathrm{N}, [\underline{\mathrm{S}}, \overline{\mathrm{S}}] \rangle, \tag{33}$$

where the second component expresses the whole range of values of the Shannon measure that are obtained for all probability measures that are consistent with the given body of evidence. This measure is more expressive than the other two in the sense that it captures all possible values of the Shannon measure that are associated with the given body of evidence. However, though the measure makes good sense conceptually, it is not yet clear how to define a meaningful ordering for it, one that would satisfy the requirement of subadditivity. Moreover, no efficient algorithm for computing $\underline{\mathrm{S}}$ has been developed as yet.

Due to the mathematically sound generalization of the measure of nonspecificity by Abellan and Moral [1] and due to the fact that the measures $\underline{\mathrm{S}}$ and $\overline{\mathrm{S}}$ can be defined for any closed convex set of probability distributions, these three measures of total uncertainty are applicable not only to DST and possibility theory, but also to a considerably more general theory based on arbitrary closed convex sets of probability distributions. This means, in turn, that they are applicable to all theories of imprecise probabilities that are subsumed under this highly general theory (as depicted in figure 2).

7. Principles of uncertainty

Although the utility of relevant uncertainty measures is as broad as the utility of any relevant measuring instrument, their role is particularly significant in three fundamental principles for managing uncertainty. These principles are a principle of minimum uncertainty, a principle of maximum uncertainty, and a principle of uncertainty invariance [20,24].

The principle of minimum uncertainty is basically an arbitration principle. It facilitates the selection of meaningful alternatives from solution sets obtained by solving problems in which some of the initial information is inevitably reduced in the solutions to various degrees. According to this principle, we should accept only those solutions in a given solution set for which the loss of the information is as small as possible. This means, in turn, that we should accept only solutions with minimum uncertainty.

Examples of problems for which the principle of minimum uncertainty is applicable are simplification problems and conflict resolution problems of various types. Consider, for example, that we want to simplify a finite-state nondeterministic system by coarsening state sets of its variables. This requires that each state set be partitioned in a meaningful way (e.g., preserving a given order of states) into a given number of subsets. This can usually be done in many different ways. The minimum uncertainty principle allows us to compare the various competing partitions by their amount of relevant uncertainty (predictive, diagnostic, etc.) and consider only those with minimum uncertainty.

As another example, let us consider a set of systems, $S_1, S_2, \ldots, S_n$, that share some variables. These systems are locally inconsistent in the sense that projections from individual systems into variables they share (e.g., marginal probabilities, marginal bodies of evidence, etc.) are not the same. To resolve the local inconsistencies, we need to replace each system S_i with another system, C_i, such that systems $C_1, C_2, \ldots, C_n$ be locally consistent. This, of course, can be done in many different ways, but the proper way to do that is to minimize the loss of information caused by these replacements. Denoting the relevant uncertainty measure by UNC, the principle of minimum uncertainty is applicable to deal with this problem by formulating the following optimization problem:

$$\text{Minimize} \sum_{i=1}^{n} \bigl[\mathrm{UNC}(C_i) - \mathrm{UNC}(S_i)\bigr]$$

subject to the following three types of constraints:

- axioms of the theory in which systems S_i and C_i $(i = 1, 2, \ldots, n)$ are formulated;
- equations by which all conditions of local consistency among systems $C_1, C_2, \ldots, C_n$ are defined;
- $\mathrm{UNC}(C_i) \geqslant \mathrm{UNC}(S_i)$ for all $i = 1, 2, \ldots, n$ to avoid introducing bias.

The second principle, *the principle of maximum uncertainty*, is essential for any problem that involves *ampliative reasoning*. This is reasoning in which conclusions are not entailed in the given premises. Using common sense, the principle may be expressed as follows: in any ampliative inference, use all information supported by available evidence but make sure that no additional information (unsupported by given evidence) is unwittingly added. Employing the connection between information and uncertainty, this definition can be reformulated in terms of uncertainty: any conclusion resulting from any ampliative inference should maximize the relevant uncertainty within constraints representing given premises. This principle guarantees that we fully recognize our ignorance when we attempt to make inferences that are beyond the information domain defined by the given premises and, at the same time, that we utilize all information contained in the premises. In other words, the principle guarantees that our inferences are maximally non-committal with respect to information that is not contained in the premises.

Let f and $\mathrm{UNC}(f)$ denote, respectively, a relevant uncertainty function (probability or possibility distribution function, basic probability assignment function, etc.) and the associated measure of uncertainty. Then, the principle of maximum uncertainty is operationally formulated in terms of the following generic optimization problem:

Determine f for which $\mathrm{UNC}(f)$ reaches its maximum under the following constraints:

- axioms upon which uncertainty functions is based;

- constraints $E_1, E_2, \ldots$, which represent partial information about f (e.g., marginal distributions, lower or upper bounds, values of f for some sets, etc.).

Ampliative reasoning is indispensable to science and engineering in many ways and, hence, the underlying principle of maximum uncertainty has great utility. For example, whenever we make predictions based on a given scientific model, we employ ampliative reasoning. Similarly, estimating microstates from the knowledge of relevant macrostates and partial knowledge of the microstates (as in image processing and many other problems) requires ampliative reasoning. The problem of the identification of an overall system from some of its subsystems is another example that involves ampliative reasoning and, hence, the principle of maximum uncertainty.

The principles of minimum and maximum uncertainty are well developed within classical information theory, where they are referred to as the principles of minimum and maximum entropy. The great utility of these principles, particularly in developing predictive models, is perhaps best demonstrated by the work of Christensen [4–6]. The principle of maximum entropy was presumably founded by Jaynes in the 1950s [17]. It has thus a history of more than 40 years. The literature concerned with this principle is extensive. The book by Kapur [18] is an excellent overview of the astonishing range of applications of the principle.

Optimization problems that emerge from the minimum and maximum uncertainty principles outside classical information theory have yet to be properly investigated and tested in praxis. Each of the three measures of total uncertainty in DST, which are introduced in section 6.5, open special challenges in this area of research. In our view, however, it is sufficient to use nonspecificity alone in most applications of the principle. By minimizing or maximizing nonspecificity, we fully control impressions in probabilities in line with given evidence. Using only nonspecificity has two advantages. First, measures of nonspecificity are well justified in all theories of imprecise probabilities. Second, measures of nonspecificity are linear functions and, hence, no methods of nonlinear optimization are needed, contrary to minimum and maximum entropy principles.

The third uncertainty principle, the principle of uncertainty invariance (also called the principle of information preservation), is of relatively recent origin [19]. It was introduced to facilitate meaningful transformations between the various uncertainty theories. According to this principle, the amount of uncertainty (and the associated uncertainty-based information) should be preserved in each transformation of uncertainty from one mathematical framework to another. Examples of applications of this principle are probability-possibility transformations and probabilistic or possibilistic approximations of general bodies of evidence in DST or, more generally, approximations of imprecise probabilities formalized in a particular theory by their counterparts in a less general theory.

Thus far, significant results have been obtained for uncertainty-invariant probability–possibility transformations. It was determined by a thorough mathematical analysis [12] that these transformations do exist and are unique only under log-interval scales. They are also meaningful, but not unique, under ordinal scales [23]. In this case, additional, context-dependent requirements may be used to make the transformations unique.

8. Conclusions

The importance of the systematic development of uncertainty theories that expand our ability to represent and manipulate additional types of uncertainty cannot be underestimated. As mentioned in the introduction, attitudes towards uncertainty are undergoing a revolution and are rapidly maturing. Whereas uncertainty was once regarded as an evil to be avoided at all costs, most researchers have (sometimes grudgingly) admitted that uncertainty often cannot be avoided and that it is important to learn to address uncertainty in a rigorous and principled way.

Beyond this reluctant acceptance of the role of uncertainty in scientific inquiry, some researchers have taken an additional step in their attitude towards uncertainty and have come to regard it as being useful and desirable. The expression of this latest attitude towards uncertainty is most clearly recognized in the field of intelligent systems, where it is manifested by the emphasis on soft computing [52]. Researchers in this field value the robust and adaptive nature of the intelligence exhibited by natural organisms, particularly that of human beings. The exhaustive brute-force approach to intelligence, typical in the current mainstream in the broader area of artificial intelligence (where a response must be programmed to every conceivable scenario) has failed to produce robustly intelligent machines.

The basic thesis of soft computing is that precision and certainty carry a cost and that intelligent systems should exploit, whenever possible, the tolerance for imprecision and uncertainty. This is exactly what human beings do when they perform complex tasks. For example, when we drive or park a car in a city, we act on the basis of perceptions and not on the basis of precise, numerical measurements. Perceptions allow us to capture the essence of each of the complex situations involved by operating at such levels of imprecision and uncertainty that are acceptable for performing the given tasks. Requiring high precision and certainty in describing these situations, as typical in traditional science and engineering, would be quite counterproductive in this case. It is easy to imagine that such a requirement would inevitably result in unmanageable complexity. To understand the many facets of uncertainty is thus crucial for designing intelligent systems.

Uncertainty also plays a central role in the adaptive intelligence of human beings. Rather than drawing upon preprogrammed responses to the conditions of our environment, human intelligence (at its best) generalizes past experience to current conditions. It is not realistic to store responses to every set of environmental conditions that we might eventually face. In order to avoid the combinatorial explosion in storage capacity required for "exhaustive" intelligence, human beings employ trial and error methods that have yet to be fully realized in machines. In this sense, understanding the role of uncertainty in intelligence and learning to strategically incorporate it into our models and designs is the new frontier of science and technology.

As figure 1 suggests, we are in our infancy with respect to a comprehensive program of formalized uncertainty languages and measures. Even where formalizations of uncertainty currently exist, our usage of them is often awkward and immature. Controversy still abounds with respect to what types of uncertainty are appropriate to express within a given language, as well as to which language is the most appropriate to express a given type of uncertainty. As discussed in section 1, a fully operational formal language of uncertainty requires development at four levels. Currently, classical numerical probability is the only theory that is fully developed at each of the four levels. As for the remaining entries in figure 1, advances will come as we progress along both dimensions of the matrix. In many cases there is progress to be made with respect to the development of the language itself (levels 1 and 2). Where formal languages have already been developed, research efforts should focus on developing measures of uncertainty-based information (i.e., level 3). Where measures of uncertainty-based information have been established, as in the theories indicated by labels 2–9 in figure 2, applications of the measures via uncertainty principles should be energetically pursued.

References

[1] J. Abellan and S. Moral, A non-specificity measure for convex sets of probability distributions, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 8(3) (2000) 357–367.

[2] A. Chateauneuf and J.Y. Jaffray, Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion, Mathematical Social Sciences 17 (1989) 263–283.

[3] G. Choquet, Theory of capacities, Annales de L'Institut Fourier 5 (1953–1954) 131–295.

[4] R. Christensen, *Entropy Minimax Sourcebook* (Entropy Limited, Lincoln, MA, 1980–1981).

[5] R. Christensen, Entropy minimax multivariate statistical modeling – I: Theory, International Journal of General Systems 11(3) (1985) 231–277.

[6] R. Christensen, Entropy minimax multivariate statistical modeling – II: Applications, International Journal of General Systems 12(3) (1986) 227–305.

[7] G. De Cooman, Possibility theory, International Journal of General Systems 25(4) (1997) 291–371.

[8] D. Dubois and H. Prade, A note on measures of specificity for fuzzy sets, International Journal of General Systems 10(4) (1985) 279–283.

[9] D. Dubois and H. Prade, *Possibility Theory* (Plenum, New York, 1988).

[10] D. Dubois and H. Prade, Rough fuzzy sets and fuzzy rough sets, International Journal of General Systems 17(2–3) (1990) 191–209.

[11] D. Dubois, J. Lang and H. Prade, Possibilistic logic, in: eds. D.M. Gabbay, et al., *Handbook of Logic in Artificial Intelligence and Logic Programming* (Clarendon, Oxford, UK, 1994), pp. 439–513.

[12] J.F. Geer and G.J. Klir, A mathematical analysis of information preserving transformations between probabilistic and possibilistic formulations of uncertainty, International Journal of General Systems 20(2) (1992) 143–176.

[13] P.R. Halmos, *Measure Theory* (Van Nostrand, Princeton, NJ, 1950).

[14] R.V.L. Hartley, Transmission of information, The Bell System Technical Journal 7(3) (1928) 535–563.

[15] M. Higashi and G.J. Klir, Measures of uncertainty and information based on possibility distributions, International Journal of General Systems 9(1) (1983) 43–58.

[16] G.E. Hughes and M.J. Cresswell, *A New Introduction to Modal Logic* (Routledge, London and New York, 1996).

[17] E.T. Jaynes, in: *Papers on Probability, Statistics and Statistical Physics*, ed. R.D. Rosenkrantz (Reidel, Dordrecht, 1983).

[18] J.N. Kapur, *Maximum Entropy Models in Science and Engineering* (Wiley, New York, 1989).

[19] G.J. Klir, A principle of uncertainty and information invariance, International Journal of General Systems 17(2–3) (1990) 249–275.

[20] G.J. Klir, Principles of uncertainty: What are they? Why do we need them?, Fuzzy Sets and Systems 74(1) (1995) 15–31.

[21] G.J. Klir, On fuzzy-set interpretation of possibility theory, Fuzzy Sets and Systems 108(3) (1999) 263–273.

[22] G.J. Klir and M. Mariano, On the uniqueness of possibilistic measure of uncertainty and information, Fuzzy Sets and Systems 24(2) (1987) 197–219.

[23] G.J. Klir and B. Parviz, Probability–possibility transformations: A comparison, International Journal of General Systems 21(3) (1992) 291–310.

[24] G.J. Klir and M.J. Wierman, *Uncertainty-Based Information: Elements of Generalized Information Theory* (Physica-Verlag/Springer, Heidelberg and New York, 1999).

[25] G.J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications* (Prentice Hall, Upper Saddle River, NJ, 1995).

[26] G.J. Klir and B. Yuan, On nonspecificity of fuzzy sets with continuous membership functions, in: *Proc. 1995 International Conf. on Systems, Man, and Cybernetics*, Vancouver (1995).

[27] G.J. Klir and B. Yuan, eds., *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi A. Zadeh* (World Scientific, Singapore, 1996).

[28] A.N. Kolmogorov, *Foundations of the Theory of Probability* (Chelsea, New York, 1950), first published in German in 1933.

[29] H.E. Kyburg, Bayesian and non-Bayesian evidential updating, Artificial Intelligence 31 (1987) 271–293.

[30] H.E. Kyburg and M. Pittarelli, Set-based Bayesianism, IEEE Transactions on Systems, Man, and Cybernetics A 26(3) (1996) 324–339.

[31] Y. Pan and G.J. Klir, Bayesian inference based on interval probabilities, Journal of Intelligent and Fuzzy Systems 5(3) (1997) 193–203.

[32] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning About Data* (Kluwer, Boston, 1991).

[33] H. Prade, Modal semantics and fuzzy set theory, in: *Fuzzy Set and Possibility Theory: Recent Developments*, ed. R.R. Yager (Pergamon Press, Oxford, 1982) pp. 232–246.

[34] A. Ramer, Uniqueness of information measure in the theory of evidence, Fuzzy Sets and Systems 24(2) (1987) 183–196.

[35] A. Ramer, Euclidean specificity: two solutions and few problems, in: *Proc. World Congress of the International Fuzzy Systems Assoc.*, Vol. 4, Prague (1997) 268–271.

[36] A. Ramer, Nonspecificity in $\mathbb{R}^n$, International Journal of General Systems (2001) in press.

[37] A. Rènyi, *Probability Theory* (North-Holland, Amsterdam, 1970) chapter IX, pp. 540–616.

[38] G. Shafer, *A Mathematical Theory of Evidence* (Princeton Univ. Press, Princeton, NJ, 1976).

[39] C.E. Shannon, The mathematical theory of communication, The Bell System Technical Journal 27 (1948) 379–423, 623–656.

[40] R.M. Smith, Generalized information theory: resolving some old questions and opening some new ones, Ph.D. dissertation, Binghamton University – SUNY, Binghamton (2000).

[41] P. Walley, *Statistical Reasoning With Imprecise Probabilities* (Chapman and Hall, London, 1991).

[42] P. Walley, Towards a unified theory of imprecise probability, International Journal of Approximate Reasoning 24(2–3) (2000) 125–148.

[43] Z. Wang and G.J. Klir, *Fuzzy Measure Theory* (Plenum, New York, 1992).

[44] K. Weichselberger and S.K. Pöhlmann, *A Methodology for Uncertainty in Knowledge-Based Systems* (Springer, New York, 1990).

[45] R.R. Yager, A foundation for a theory of possibility, Journal of Cybernetics 10(1–3) (1980) 77–204.

[46] R.R. Yager, On the Dempster–Shafer framework and new combination rules, Information Sciences 41 (1987) 93–137.

[47] R.R. Yager et al., *Fuzzy Sets and Applications – Selected Papers by L.A. Zadeh* (Wiley, New York, 1987).

[48] J. Yen, Generalizing the Dempster–Shafer theory to fuzzy sets, IEEE Transactions on Systems, Man, and Cybernetics 20(3) (1990) 559–570.
[49] L.A. Zadeh, Fuzzy sets, Information and Control 8(3) (1965) 338–353.
[50] L.A. Zadeh, Probability measures of fuzzy events, Journal of Mathematical Analysis and Applications 23 (1968) 421–427.
[51] L.A. Zadeh, Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets and Systems 1(1) (1978) 3–28.
[52] L.A. Zadeh, Soft computing and fuzzy logic, IEEE Software 11(6) (1994) 48–56.