POSSIBILISTIC INFORMATION: A Tutorial

George J. Klir
State University of New York (SUNY)
Binghamton, New York 13902, USA
gklir@binghamton.edu

prepared for International Centre for Information and Uncertainty, Palacky University, Olomouc

european social fund in the czech republic
EUROPEAN UNION
MINISTRY OF EDUCATION, YOUTH AND SPORTS
OP Education for Competitiveness

INVESTMENTS IN EDUCATION DEVELOPMENT

OUTLINE

1. Generalized information theory (GIT)
2. Uncertainty theories
3. Uncertainty-based information
4. Uncertainty-based information in systems
5. Classical possibilistic uncertainty theory
6. Classical possibilistic functional *U* for measuring the amount of uncertainty (Hartley measure)
7. Uncertainty theory with graded possibilities
8. Theory of random sets
9. Summary of results and some open problems

GENERALIZED INFORMATION THEORY (GIT)

- GIT is a research program whose objective is to develop a formal treatment of the dual concepts of uncertainty and information in all their varieties.
- GIT was introduced in my paper published in 1991 in Fuzzy Sets and Systems, 40(1), pp. 127-142.
- In GIT, as in the two classical information theories (possibilistic and probabilistic), uncertainty is the primary concept and information is defined as the capacity to reduce uncertainty.

UNCERTAINTY THEORIES:
General Scenario

- A set of mutually exclusive alternatives is considered (predictions, diagnoses, etc.).
- Only one of the alternatives is true, but we are not certain which one it is.
- Uncertainty is expressed differently in each theory.

CLASSICAL UNCERTAINTY THEORIES

- **POSSIBILISTIC:** Uncertainty results from more possible alternatives than one. Information is obtained by any evidence that some of the considered alternatives are not possible.

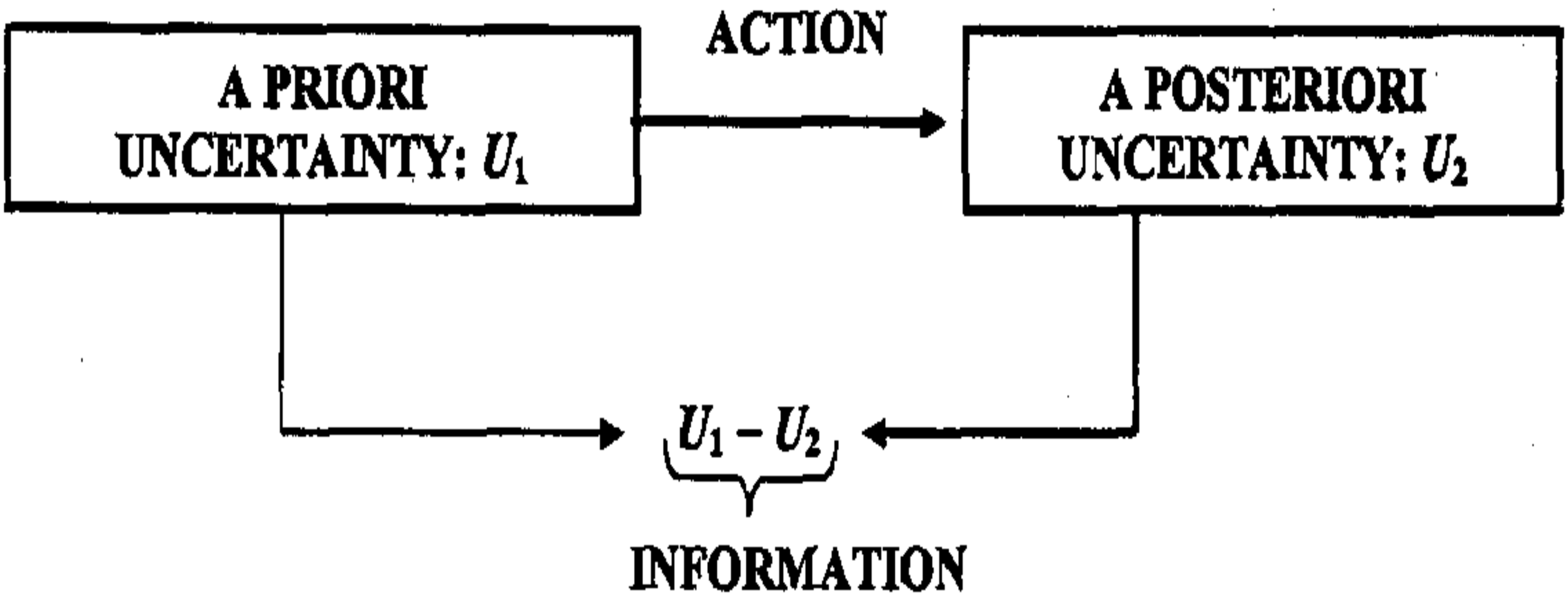
- **PROBABILISTIC:** Uncertainty results from a distribution of degrees of evidential claims from a fixed value among all considered alternatives. Information is obtained by any evidence that makes the distribution more discriminatory.

UNCERTAINTY THEORIES: Levels of Development

1. Uncertainty in theory T is formalized in terms of a class of appropriate uncertainty functions u.
2. Operating rules (calculus) for manipulating functions u in theory T are developed.
3. A justifiable functional, U, is found that for each particular uncertainty function u in theory T measures the amount of uncertainty captured by u.
4. Methodological aspects of theory T are developed, including the use of functional U as an abstract measuring instrument.

UNCERTAINTY-BASED INFORMATION

The amount of information obtained by an action
is equal to
the amount of uncertainty reduced by the action.


A PRIORI
UNCERTAINTY: U_1
ACTION
A POSTERIORI
UNCERTAINTY: U_2
$\underbrace{U_1 - U_2}$
INFORMATION

INFORMATION IN SYSTEMS

- Every system is constructed within a chosen experimental frame for some purpose.
- The system is ultimately a formal description of a constraint among variables of the experimental frame.
- The constraint is utilized in a purposeful way for restricting states of some variables on the basis of known states of other variables.
- Every system can be asked various relevant questions and information contained in each answer obtained by the system can be measured by reduction of its uncertainty with respect to the associated experimental frame.

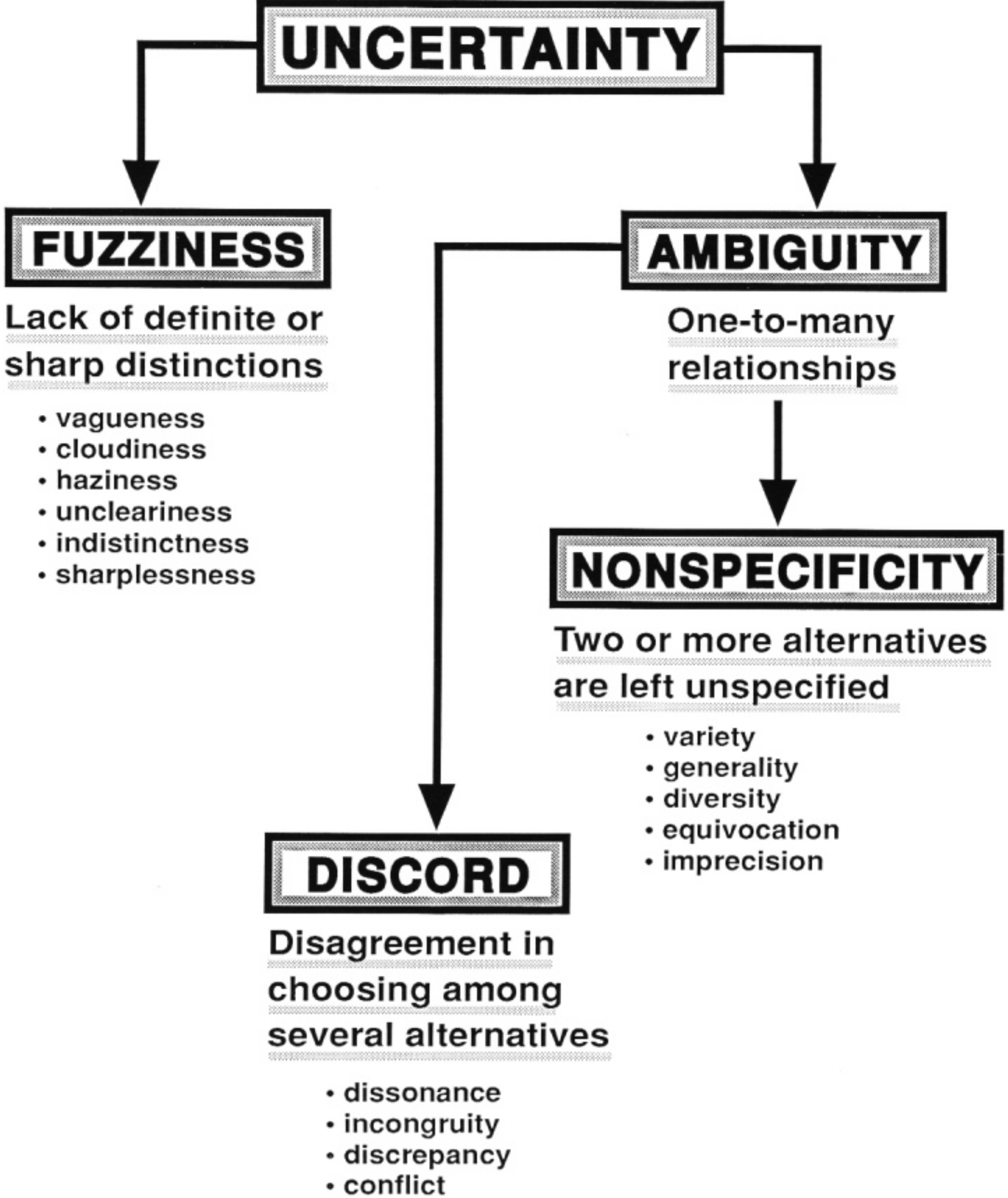
EXPERIMENTAL FRAME

- A set of variables chosen on an object of interest for some purpose (prediction, retrodiction, etc.).
- A set of states that are recognized for each chosen variable (numerical or non-numerical).
- Supporting media within which the variables change their states (time, space, population).

INFORMATION IN SYSTEMS

- The amount of information in a given system with respect to a particular question = the amount of uncertainty contained in the answer obtained solely within the experimental frame (in face of total ignorance) – the amount of uncertainty contained in the answer by using the system.

- INFORMATION(system, question) = UNCERTAINTY(exp. frame, question) – UNCERTAINTY(system,question)


UNCERTAINTY
FUZZINESS
Lack of definite or sharp distinctions
• vagueness
• cloudiness
• haziness
• uncleariness
• indistinctness
• sharplessness
AMBIGUITY
One-to-many relationships
NONSPECIFICITY
Two or more alternatives are left unspecified
• variety
• generality
• diversity
• equivocation
• imprecision
DISCORD
Disagreement in choosing among several alternatives
• dissonance
• incongruity
• discrepancy
• conflict

VAGUENESS -- FUZZINESS

A proposition is vague when there are possible states of things concerning which it is *intrinsically uncertain* whether, had they been contemplated by the speaker, he would have regarded them as excluded or allowed by the proposition. By intrinsically uncertain we mean not uncertain in consequence of any ignorance of the interpreter, but because the speaker's habits were indeterminate.

Charles S. Peirce

NONSPECIFITY

The formal condition for the presence of uncertainty in a person's thought is the presence there of a question to which he cannot exclude all except one answer.

(George Shackle, 1979, p. 147)

POSSIBILISTIC UNCERTAINTY

When you have eliminated the impossible,
whatever remains must be the case,
however improbable it may seem to be.

(Sherlock Holmes)

PROBABILISTIC UNCERTAINTY

Probability is degree of certainty and differs from absolute certainty as the part differs from the whole.

(Jacques Bernouli)

CLASSICAL POSSIBILISTIC UNCERTAINTY THEORY

- Given a finite set X of considered alternatives, uncertainty is expressed by a possibility function

$$r: X \rightarrow \{0,1\}.$$

- $r(x) = 1$ means that x is possible and $r(x) = 0$ means that, under given evidence, x is not possible, $x \in X$.
- Function r partitions set X into two subsets: X_0 and X_1.
- Information is obtained by any evidence that reduces the subset X_1 of possible alternatives.
- Possibility set function (nonadditive measure):

$$\text{Pos}(A) = \max_{x \in A}\{r(x)\}, \forall A \subseteq X$$

AXIOMS OF CLASSICAL POSSIBILITY THEORY

1. Pos $(\varnothing) = 0$;

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

3. For any A, $B \subseteq X$, $\mathrm{Pos}(A \cup B) = \max\{\mathrm{Pos(A)}, \mathrm{Pos(B)}\}$.

DUALITY BETWEEN POSSIBILITY AND NECESSITY MEASURES

For each possibility measure Pos, a dual necessity measure, Nec, is defined by the equation

$$\mathrm{Nec}(A) = 1 - \mathrm{Pos}(\bar{A}) \text{ for all } A \subseteq X.$$

It is necessary that the true element is in A if it is not possible that it is in the complement of A.

SOME BASIC PROPERTIES OF POSSIBILITY AND NECESSITY MEASURES

- $\mathrm{Nec}(A) \leq \mathrm{Pos}(A)$
- $\mathrm{Pos}(A \cap B) \leq \min[\mathrm{Pos}(A), \mathrm{Pos}(B)]$
- $\mathrm{Nec}(A \cup B) \geq \max[\mathrm{Nec}(A), \mathrm{Nec}(B)]$
- $\mathrm{Pos}(A) + \mathrm{Pos}(\bar{A}) \geq 1$
- $\mathrm{Nec}(A) + \mathrm{Nec}(\bar{A}) \leq 1$
- $\max[\mathrm{Pos}(A), \mathrm{Pos}(\bar{A})] = 1$
- $\min[\mathrm{Nec}(A), \mathrm{Nec}(\bar{A})] = 0$
- $\mathrm{Pos}(A) < 1 \Rightarrow \mathrm{Nec}(A) = 0$
- $\mathrm{Nec}(A) > 0 \Rightarrow \mathrm{Pos}(A) = 1$

FUNCTIONALS *U* FOR MEASURING UNCERTAINTY: Key Requirements

1. Subadditivity
2. Additivity
3. Range
4. Continuity
5. Normalization
6. Expansibility
7. Consistency
8. Monotonicity
9. Coordinate Invariance

UNIQUENESS OF FUNCTIONAL U

- For any given uncertainty theory T, functional U is required to be unique under all the requirements formulated in the calculus of theory T.
- The normalization requirement may be formulated in different ways in each theory T. Each possible formulation defines a particular measurement unit of uncertainty when uncertainty is measured by functional U in theory T.
- U is an abstract measuring instrument: function u in theory T is an input, a real number measuring the amount of uncertainty captured by u is the corresponding output.

RÉNYI' S AXIOMS FOR MEASURE OF POSSIBILISTIC UNCERTAINTY U

$$U\colon |X_1| \rightarrow [0, \infty)$$

Axiom 1. $U(n \times m) = U(n) + U(m)$ [additivity].

Axiom 2. $U(n) \leq U(n+1)$ [monotonicity].

Axiom 3. $U(2) = 1$[normalization].

UNIQUENESS OF POSSIBILISTIC MEASURE OF UNCERTAINTY U

Theorem (Rényi 1970). $U(X_1) = \log_2 |X_1|$ is the only functional that satisfies Axioms 1-3.

Proof: **Part 1**

1. For any integer $n \geq 2$ and any natural number i, define a natural number $q(i)$ such that $2^{q(i)} \leq n^i < 2^{q(i)+1}$.
2. Rewrite as: $q(i)\log_2 2 \leq i \log_2 n < [q(i)+1]\log_2 2$.
3. Rewrite again as: $q(i)/i \leq \log_2 n < (q(i)+1)/i$.
4. Hence, $\lim_{i\to\infty}(q(i)/i) = \log_2 n$.

Part 2

1. By monotonicity, $U(2^{q(i)}) \leq U(n^i) \leq U(2^{q(i)+1})$.
2. By additivity, $U(a^k) = U(a^{k-1})+U(a) = U(a^{k-2})+U(a)+U(a) = U(a^{k-3})+U(a)+U(a)+U(a) = \ldots = kU(a)$.
3. Hence, $q(i)U(2) \leq iU(n) \leq (q(i) + 1)U(2)$.
4. By normalization, $q(i) \leq iU(n) \leq q(i) + 1$.
5. Rewrite as, $q(i)/i \leq U(n) \leq (q(i)+1)/i$.
6. Hence, $\lim_{i\to\infty}(q(i)/i) = U(n)$.
7. From Part 1 of the proof: $\lim_{i\to\infty}(q(i)/i) = \log_2 n$.
8. Hence, $U(n) = \log_2 n$.
9. Than, if $n = |X_1|$, $U(|X_1|) = \log_2|X_1|$.

Hartley functional for measuring classical possibilistic uncertainty, which is usually referred to as nonspecificity, and the associated uncertainty-based information:

$$H(A) = \log_2 |A|$$

$$I_H(A) = \log_2 |X| - \log_2 |A|$$

JOINT AND MARGINAL HARTLEY MEASURES ON $X \times Y$

- **Joint Hartley measure**: $H(X,Y) = \log_2|R|$, where R is a relation on $X \times Y$ (i.e. $R \subseteq X \times Y$).

- **Marginal Hartley measures**: $H(X) = \log_2|R_X|$ and $H(Y) = \log_2|R_Y|$, where
  $R_X = \{x \in X: (x,y) \in R \text{ for some } y \in Y\}$,
  $R_Y = \{y \in X: (x,y) \in R \text{ for some } x \in X\}$.

CONDITIONAL HARTLEY MEASURE ON $X \times Y$

- $H(X \mid y) = \log_2|X_y|$, where $X_y = \{x \in X: (x, y) \in R$ for a particular $y \in Y$, and $\sum_{y \in Y} |X_y| = R$.
- $H(X \mid Y) = \log_2|R|/|R_Y| = \log_2|R| - \log_2|R_Y| = H(X,Y) - H(Y)$.
- Observe that $|R|/|R_Y|$ is the average, for all $y \in Y$, of numbers of possible alternatives in X when individual alternatives in Y are chosen.
- Similarly, $H(Y \mid X) = H(X,Y) - H(X)$.

Basic Equations and Inequalities of Uncertainty Measures U *on* $X \times Y$

- $U(X\,|Y) = U(X \times Y) - U(Y)$
- $U(Y\,|X) = U(X \times Y) - U(X)$
- $T_U(X, Y) = U(X) + U(Y) - U(X \times Y)$
- $U(X \times Y) \leq U(X) + U(Y)$
- $U(X\,|Y) \leq U(X)$ and $U(Y\,|X) \leq U(Y)$
- These equations and inequalities are valid in all theories of uncertainty.

HARTLEY-LIKE MEASURE IN *n*-DIMENSIONAL EUCLIDEAN SPACE

$$HL(A) = \min_{t \in T} \left\{ c \log_2 \left[\prod_{i=1}^{n} \left[1 + \mu\left(A_{i_t}\right) \right] + \mu(A) - \prod_{i=1}^{n} \mu\left(A_{i_t}\right) \right] \right\}$$

- A is a convex subset of $\mathbf{R}^n$.
- μ denotes the Lebesgue measure.
- T denotes the set of all isometric transformations from one orhogonal coordinate system to another.
- A_{i_t} denotes the i-th projection of A within the coordinate system t.

THEORY OF GRADED POSSIBILITIES

- Classical possibility function $r: X \rightarrow \{0,1\}$ is in this theory generalized to $r: X \rightarrow [0,1]$.
- $r(x)$ is interpreted as the degree to which x is possible.
- Possibility measure, Pos, is defined by the same axioms as in classical possibility theory.
- A dual necessity measure, Nec, is again defined by

$$\mathrm{Nec}(A) = 1 - \mathrm{Pos}(\bar{A}), \ \forall A \subseteq X.$$

JOINT AND MARGINAL GRADED POSSIBILITIES ON $X \times Y$

Given a joint possibility function on $r: X \times Y \rightarrow [0,1]$, the associated marginal functions are defined by:

$$r_X(x) = \max_{y \in Y}\{r\{x, y\}, \forall x \in X,$$

$$r_Y(y) = \max_{x \in X}\{r(x, y\}, \forall y \in Y.$$

POSSIBILISTIC INDEPENDENCE ON $X \times Y$

- In general, $r(x,y) \leq \min\{r_X(x), r_Y(y)\}$ for all $(x,y) \in X \times Y$.
- Possiblistic independence is obtained when $r(x,y) = \min\{r_X(x), r_Y(y)\}$.

CONDITIONAL GRADED POSSIBILITIES ON $X \times Y$

In general, the joint possibility function can be expressed via either of the following equations:

$$r(x,y) = \min\left\{r_Y(y), r_{X|Y}(x|y)\right\},$$

$$r(x,y) = \min\left\{r_X(x), r_{Y|X}(y|x)\right\}.$$

CONDITIONAL GRADED POSSIBILITIES ON $X \times Y$ -- 2

By solving the equations for $r_{X|Y}$ and $r_{Y|X}$, we obtain:

$$r_{X|Y}(x|y) = \begin{cases} [r(x,y),1] & \text{when } r_Y(y) = r(x,y) \\ r(x,y) & \text{when } r_Y(y) > r(x,y) \end{cases}$$

$$r_{Y|X}(y|x) = \begin{cases} [r(x,y),1] & \text{when } r_X(x) = r(x,y) \\ r(x,y) & \text{when } r_X(x) > r(x,y) \end{cases}$$

ORDERING OF POSSIBILITY DEGREES

- Let $X = \{x_1, x_2, \ldots, x_n\}$ and let $r(x_i) \geq r(x_{i+1})$ for all $i = 1, 2, \ldots, n-1$.
- Let $r_i = r(x_i)$ and let the n-tuple $\mathbf{r} = (r_1, r_2, \ldots, r_n)$, where $r_1 = 1$, be called a possibility profile.
- Standard partial ordering of possibility profiles on X forms a lattice whose maximum and minimum elements are, respectively, $(1,1, \ldots, 1)$ and $(1,0,0, \ldots, 0)$.
- For any two possibility profiles ${}^j\mathbf{r}$ and ${}^k\mathbf{r}$ such that ${}^j\mathbf{r} \leq {}^k\mathbf{r}$, ${}^j\mathbf{r}$ contains more information.

NECESSITY MEASURE

- Let $A_i = \{x_1, x_2, \ldots, x_i\}$ for all $i = 1, 2, \ldots, n$.
- Sets A_i are the only subsets of X for which the necessity measure, Nec, is not zero.
- Assuming that $r_{n+1} = 0$ by convention,we have

$$\mathrm{Nec}(A_i) = 1 - r_{i+1}.$$

MÖBIUS TRANSFORM

$$m(A) = \sum_{B:B\subseteq A} (-1)^{|A-B|}\mathrm{Nec}(B)$$

$$\mathrm{Nec}(A) = \sum_{B:B\subseteq A} m(B)$$

MÖBIUS REPRESENTATIONS

Möbius representations are set functions, *m*, that satisfy the following two requirements:

$$M(\varnothing) = 0; \text{ and}$$

$$\sum_{A \subseteq X} m(A) = 1.$$

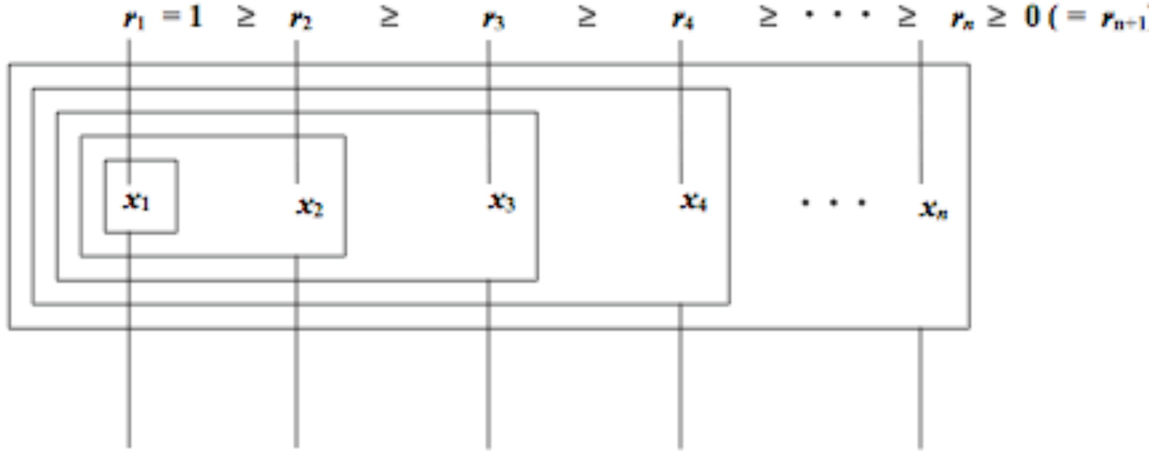
MÖBIUS REPRESENTATION IN POSSIBILITY THEORY

- Applying the Möbius transformation to a necessity measure, we obtain $m(A) = r_i - r_{i+1}$ when $A = A_i$ and $m(A) = 0$ for all other subsets of X.
- Let $m_i = m(A_i)$. Then, $m_i = r_i - r_{i+1}$ ($r_{n+1} = 0$).
- For each possibility profile **r**, there exists a unique Möbius n-tuple $\mathbf{m} = (m_1, m_2, \ldots, m_n)$ such that

$$\sum_{i=1}^{n} m_i = 1.$$

POSSIBILITY PROFILES VERSUS THE ASSOCIATED MÖBIUS TUPLES

- Information ordering of Möbius n-tuples may be induced from the natural information ordering of possibility profiles via the isomorphic relation between the two representations.
- $\mathbf{r} = (1, 1, \ldots, 1)$ corresponds to $\mathbf{m} = (0, 0, \ldots, 0, 1)$.
- $\mathbf{r} = (1, 0, \ldots, 0)$ corresponds to $\mathbf{m} = (1, 0, \ldots, 0)$.



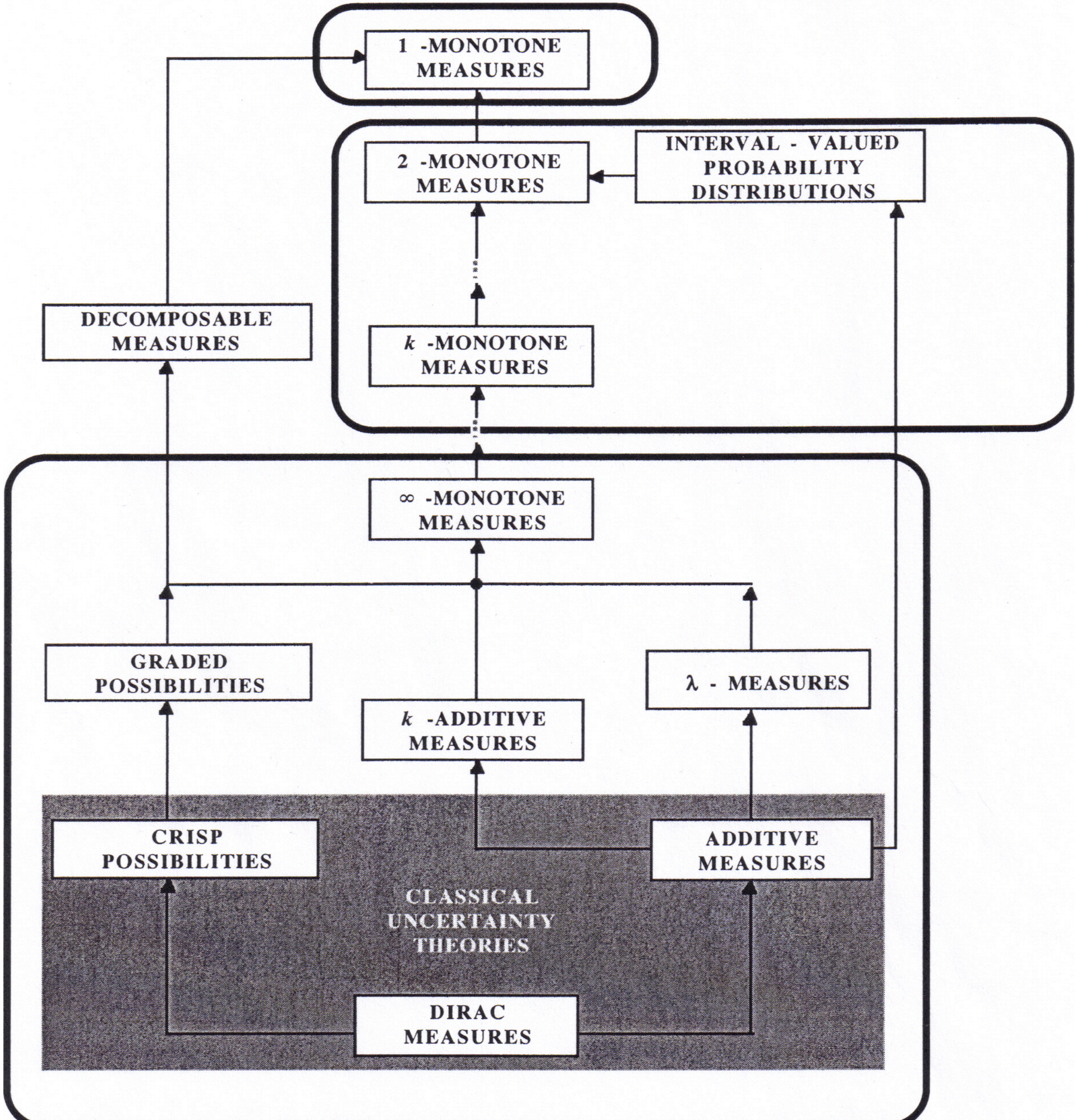
$(A_0 = \varnothing)$	A_1	A_2	A_3	A_4	$\cdots$	A_n
$m_i = m(A_i)$:	$1 - r_2$	$r_2 - r_3$	$r_3 - r_4$	$r_4 - r_5$	$\cdots$	r_n
$Nec(A_i)$:	$1 - r_2$	$1 - r_3$	$1 - r_4$	$1 - r_5$	$\cdots$	1
$Pos(A_i)$:	1	1	1	1	$\cdots$	1
$Pos(A_n - A_{i-1})$:	1	r_2	r_3	r_4	$\cdots$	r_n

GENERALIZED HARTLEY MEASURE OF NONSPECIFICITY IN THE THEORY OF GRADED POSSIBILITIES

- Introduced in (Higashi and Klir, 1983):

$$GH(\mathbf{r}) = \sum_{i=1}^{n} m_i \log_2 |A_i| = \sum_{i=1}^{n} m_i \log_2 i.$$

- Uniqueness of *GH* proved in (Klir and Mariano, 1987).


1 -MONOTONE MEASURES
2 -MONOTONE MEASURES
INTERVAL - VALUED PROBABILITY DISTRIBUTIONS
DECOMPOSABLE MEASURES
k -MONOTONE MEASURES
∞ -MONOTONE MEASURES
GRADED POSSIBILITIES
k -ADDITIVE MEASURES
λ - MEASURES
CRISP POSSIBILITIES
ADDITIVE MEASURES
CLASSICAL UNCERTAINTY THEORIES
DIRAC MEASURES

THEORY OF FINITE RANDOM SETS

- Given a finite universe X, a finite random set is a nonempty family **F** of subsets of X and a function m on **F.**
- For all $A \in \mathbf{F}$, $m(A) \in [0,1]$, $m(\varnothing)=0$, and $\sum_{A \in \mathbf{F}} m(A) = 1$.
- Sets in **F** can be viewed as sets of possible alternatives associated with values of function m.
- Function m can be viewed as a probability distribution function on sets in **F.**

RANDOM-SET THEORY AND EVIDENCE THEORY ON FINITE SETS

- Function m is a special Möbius representation (since $m(A) \in [0,1]$ for all $A \in \mathbf{F}$).
- The inverse Möbius transform results in a totally monotone measure *Bel* (belief measures) of evidence theory:

$$Bel(A) = \sum_{B:B \subseteq A} m(A).$$

- A dual measure is a plausibility measure, *Pl*, defined in the usual way: $Pl(A) = 1 - Bel(\bar{A})$.

HARTLEY MEASURE OF NONSPECIFICITY IN THE THEORY OF RANDOM SETS

- The theory of random sets is closely connected with the Dempster-Shafer theory of evidence. Both are based on totally monotone measures, for which the Möbius representation is always positive.
- The generalized Hartley measure has the form:

$$GH(m) = \sum_{A \subseteq X} m(A) \log_2 |A|.$$

MEASURES OF NONSPECIFICITY ON FINITE SETS: A Historical Overview

- In classical possibility theory:

$$H(A) = \log_2|A|.$$

- In the theory of graded possibilities:

$$GH(\mathbf{r}) = \sum_{i=1}^{n} m_i \log_2 |A_i| = \sum_{i=1}^{n} m_i \log_2 i.$$

- In the theory of random sets (or evidence theory):

$$GH(m) = \sum_{A \in \mathbf{F}} m(A) \log_2 |A.|$$

HARTLEY-LIKE MEASURE IN THE THEORY OF GRADED POSSIBILITIES

- When A is a classical set on $\mathbf{R}^n$ ($n \geq 1$), then

$$HL(A) = \min_{t \in T} \left\{ c \log_2 \left[\prod_{i=1}^{n} \left[1 + \mu\left(A_{i_t}\right) \right] + \mu(A) - \prod_{i=1}^{n} \mu\left(A_{i_t}\right) \right] \right\}$$

- When A is a standard fuzzy set on $\mathbf{R}^n$, then

$$HL(A) = \int_0^1 HL(A_\alpha)\, d\alpha.$$

SOME INTERPRETATIONS OF GRADED POSSIBILITIES

- Fuzzy-set interpretation:
$$r_F(x) = F(x) + 1 - \max_{x \in X} F(x), \forall x \in X.$$
- Similarity interpretation: $r(x)$ is the degree of similarity between x and an ideal prototype x_i.
- Comparative interpretation: $r(x)$ is, for example, the degree of ease to achieve x.
- Frequency interpretation: Given a nested family $\{A_i: i = 1,2, \ldots, n\}$ of subsets of X, $m_i = r_i - r_{i+1}$ is defined as the number of observations in A_i and
$$r_i = \sum_{k=i}^{n} m_k .$$

CONCLUSIONS AND OPEN QUESTIONS

- Theory of possibilities, both classical and graded, is now well developed at all levels.
- Algorithmic research is needed for computing the Hartley-like measure.
- The important class of decomposable measure has not been investigated from the standpoint of GIT as yet.
- Research regarding the construction of possibility profiles is still needed for some interpretations.