January 20, 2005

(To appear in Information Sciences)

Toward a Generalized Theory of Uncertainty (GTU)—An Outline

Lotfi A. Zadeh[a*1]

[a] Professor in the Graduate School and Director, Berkeley initiative in Soft Computing (BISC), Computer Science Division and the Electronics Research Laboratory, Department of EECS, University of California, Berkeley, CA 94720-1776, USA

Abstract

It is a deep-seated tradition in science to view uncertainty as a province of probability theory. The Generalized Theory of Uncertainty (GTU) which is outlined in this paper breaks with this tradition and views uncertainty in a broader perspective.

Uncertainty is an attribute of information. A fundamental premise of GTU is that information, whatever its form, may be represented as what is called a generalized constraint. The concept of a generalized constraint is the centerpiece of GTU. In GTU, a probabilistic constraint is viewed as a special—albeit important—instance of a generalized constraint.

A generalized constraint is a constraint of the form X isr R, where X is the constrained variable, R is a constraining relation, generally non-bivalent, and r is an indexing variable which identifies the modality of the constraint, that is, its semantics. The principal constraints are: possibilistic (r=blank); probabilistic ($r=p$); veristic ($r=v$); usuality ($r=u$); random set ($r=rs$); fuzzy graph ($r=fg$); bimodal ($r=bm$); and group ($r=g$). Generalized constraints may be qualified, combined and propagated. The set of all generalized constraints together with rules governing qualification, combination and propagation constitutes the Generalized Constraint Language (GCL).

The Generalized Constraint Language plays a key role in GTU by serving as a precisiation language for propositions, commands and questions expressed in a natural language. Thus, in GTU the meaning of a proposition drawn from a natural language is expressed as a generalized constraint. Furthermore, a proposition plays the role of a carrier of information. This is the basis for equating information to a generalized constraint

In GTU, reasoning under uncertainty is treated as propagation of generalized constraints, in the sense that rules of deduction are equated to rules which govern propagation of generalized constraints. A concept which plays a key role in deduction is

[*] Telephone: +1-510-642-4959; Fax: +1-510-642-1712.
E-mail: zadeh@eecs.berkeley.edu .

[1] Dedicated to Didier Dubois, Henri Prade and the memory of my mentors, Richard Bellman and Herbert Robbins.

that of a protoform (abbreviation of prototypical form). Basically, a protoform is an abstracted summary—a summary which serves to identify the deep semantic structure of the object to which it applies. A deduction role has two parts: symbolic—expressed in terms of protoforms—and computational.

GTU represents a significant change both in perspective and direction in dealing with uncertainty and information. The concepts and techniques introduced in this paper are illustrated by a number of examples.

1. Introduction

Uncertainty is an attribute of information. The path-breaking work of Shannon has led to a universal acceptance of the thesis that information is statistical in nature. A logical consequence of this thesis is that uncertainty, whatever its form, should be dealt with through the use of probability theory. To quote an eminent Bayesian, Professor Dennis Lindley, "The only satisfactory description of uncertainty is probability. By this I mean that every uncertainty statement must be in the form of a probability; that several uncertainties must be combined using the rules of probability; and that the calculus of probabilities is adequate to handle all situations involving uncertainty…probability is the only sensible description of uncertainty and is adequate for all problems involving uncertainty. All other methods are inadequate…anything that can be done with fuzzy logic, belief functions, upper and lower probabilities, or any other alternative to probability can better be done with probability," (Lindley, 1987).

The Generalized Theory of Uncertainty (GTU) is a challenge to the thesis and its logical consequence. Basically, GTU puts aside the thesis and its logical consequence, and adopts a much more general conceptual structure in which statistical information is just one—albeit an important one—of many forms of information. More specifically, the principal premise of GTU is that, fundamentally, information is a generalized constraint on the values which a variable is allowed to take. The centerpiece of GTU is the concept of a generalized constraint—a concept drawn from fuzzy logic, as will be described in greater detail in the sequel. The distinguishing feature of fuzzy logic is that in fuzzy logic everything is—or is allowed to be—a matter of degree. The principal tools which GTU draws from fuzzy logic include Precisiated Natural Language (PNL) and Protoform Theory (PFT), (Zadeh [56]).

In GTU, uncertainty is linked to information through the concept of granular structure—a concept which plays a key role in human interaction with the real world, Zadeh [43, 52].

Informally, a granule of a variable *X* is a clump of values of *X* which are drawn together by indistinguishability, equivalence, similarity, proximity or functionality. For example, an interval is a granule. So is a fuzzy interval. And so is a probability distribution.

Granulation is pervasive in human cognition. For example, the granules of Age are fuzzy sets labeled young, middle-aged and old, Fig. 1. The granules of Height may be very short, short, medium, tall, and very tall. And the granules of Truth may be not true, quite true, not very true, very true, etc. The concept of granularity underlies the concept of a linguistic variable—a concept which was introduced in my 1973 paper "Outline of A New Approach to the Analysis of Complex Systems and Decision Processes," Zadeh [41, 42]. The concept of a linguistic variable plays a pivotal role in almost all applications of fuzzy logic [12], [15], [18], [29], [31], [38].

There are four basic rationales which underlie granulation of attributes and the concomitant use of linguistic variables. First, the bounded ability of sensory organs, and ultimately the brain, to resolve detail and store information. For example, looking at Monika, I see that she is young but cannot pinpoint her age as a single number. Second, when numerical information may not be available. For example, I may not know exactly how many Spanish restaurants there are in San Francisco, but my perception may be "not many." Third, when an attribute is not quantifiable. For example, we describe degrees of Honesty as: low, not high, high, very high, etc because we do not have a numerical scale. And fourth, when there is a tolerance for imprecision which can be exploited through granulation to achieve tractability, robustness and economy of communication. For example, it may be sufficient to know that Monika is young; her exact age may be unimportant. What should be noted is that this is the principal rationale which underlies the extensive use of granulation, in the form of linguistic variables, in consumer products.

There is a close connection between granularity and uncertainty. Assume that X is a variable and I am asked, "What is the value of X?" If my answer is "X is a," where a is a singleton, then there is no uncertainty in the information which I am providing about X. In this instance, information is singular. But if the answer is "X is approximately a," or "X is $*a$," for short, then there is some uncertainty in my answer. In this instance, information and its uncertainty will be described as granular. Closely, but not exactly, granularity may be equated to non-singularity. In the instance of "X is $*a$," information is non-singular.

A basic question which arises is: How can the meaning of $*a$ be precisiated? In the context of standard probability theory, call it PT, $*a$ would normally be interpreted as a probability distribution centering on a. In GTU, information about X is viewed as a generalized constraint on X. More specifically, in the context of GTU, $*a$ would be viewed as a granule which is characterized by a generalized constraint. As will be seen in the sequel, a probability distribution is a special case of a generalized constraint. In this sense, GTU is more general than PT. Actually, GTU is a generalized theory of uncertainty in the sense that most, and possibly all, existing approaches to representation of uncertain information fit within its conceptual structure. (Bloch et al [3], Bouchon-Meunier, Yager and Zadeh (eds) [4], Bubnicki [5], Dubois and Prade [8], [9], Klir [20], Shafer [33], Singpurwalla and Booker [34], Smets [35], Yager [36])

So what is the rationale for GTU? There is a demonstrable need for GTU because existing approaches to representation of uncertain information are inadequate for dealing with problems in which uncertain information is perception-based and is expressed in a natural language. (Zadeh [54]) Fig. 2. More specifically, the existing approaches do not address the problem of semantics of natural languages, Novak, Perfilieva and Mockor [25], and the need for a variety of calculi of generalized constraints to deal with it. The simple examples which follow are intended to serve as illustrations.

The Robert example. Usually Robert returns from work at about 6:00 pm. What is the probability that Robert is home at about 6:15 pm?

The balls-in-box example. A box contains about twenty black and white balls. Most are black. There are several times as many black balls as white balls. What is the number of white balls? What is the probability that a ball drawn at random is white?

The tall Swedes problem. Most Swedes are tall. What is the average height of Swedes? How many Swedes are short?

The partial existence problem. *X* is a real-valued variable; *a* and *b* are real numbers, with $a < b$. I am uncertain about the value of *X*. What I know about *X* is that (a) *X* is much larger than approximately *a*, **a*; and (b) that *X* is much smaller than approximately *b*, **b*. What is the value of *X*?

Vera's age problem. Vera has a son who is in mid-twenties, and a daughter, who is in mid-thirties. What is Vera's age?

A common thread which runs through these examples relates to the nature of given information. More specifically, the given information, e.g., "Most Swedes are tall," is perception-based and imprecise, Fig. 2. One of the basic limitations of standard probability theory, PT, is rooted in the fact that its conceptual structure does not accommodate perception-based information which is imprecise (Zadeh [55]).

To deal effectively with problems of this kind what is needed is the machinery of fuzzy logic. One of the principal tools in this machinery is granular computing, Zadeh [4], [52], [53], Bargiela and Pedrycz [1], and Lin [21]. A key concept in granular computing is that of a generalized constraint Zadeh [49]. This is why the concept of a generalized constraint is the centerpiece of GTU. A brief discussion of this concept follows.

Note: GTU draws on many concepts and techniques which relate to fuzzy logic. To facilitate understanding of GTU by those who are not conversant with fuzzy logic, our exposition includes a larger than usual number of examples and figures.

2. The Concept of a Generalized Constraint

Constraints are ubiquitous. A typical constraint is an expression of the form $X \varepsilon C$, where X is the constrained variable, and C is the set of values which X is allowed to take. A typical constraint is hard (inelastic) in the sense that if u is a value of X then u satisfies the constraint if and only if $u \varepsilon C$.

The problem with hard constraints is that most real-world constraints are not hard, that is, have some degree of elasticity. For example, the constraints "check-out time is 1 pm," and "speed limit is 100 km/hr," have, in reality, some elasticity. How can such constraints be defined? The concept of a generalized constraint is motivated by questions of this kind.

Real-world constraints may assume a variety of forms. They may be simple in appearance and yet have a complex structure. Reflecting this reality, a generalized constraint, GC, is defined as an expression of the form. (Zadeh [49]),

GC: X isr R,

where X is the constrained variable; R is a constraining relation which, in general, is non-bivalent; and r is an indexing variable which identifies the modality of the constraint, that is, its semantics.

The constrained variable, X, may assume a variety of forms. In particular,

- X is an n-ary variable, $X=(X_1, \ldots, X_n)$
- X is a proposition, e.g., X=Leslie is tall
- X is a function
- X is a function of another variable, $X=f(Y)$
- X is conditioned on another variable, X/Y
- X has a structure, e.g., X=Location(Residence(Carol))
- X is a group variable. In this case, there is a group, $G[A]$; with each member of the group, Name_i, i=1, …, n, associated with an attribute–value, A_i. A_i may be vector-valued. Symbolically

  $G[A]$: $\text{Name}_1/A_1+\ldots+\text{Name}_n/A_n$.

  Basically, $G[A]$ is a relation.

- X is a generalized constraint, $X= Y$ isr R.

A generalized constraint, GC, is associated with a test-score function, $ts(u)$, (Zadeh [45]) which associates with each object, u, to which the constraint is applicable, the degree to which u satisfies the constraint. Usually, $ts(u)$ is a point in the unit interval. However, if necessary, the test-score may be a vector, an element of a semi-ring

(Rossi [32]), an element of a lattice (Goguen [16]) or, more generally, an element of a partially ordered set, or a bimodal distribution—a constraint which will be described later in this section. The test-score function defines the semantics of the constraint with which it is associated.

The constraining relation, *R*, is, or is allowed to be, non-bivalent (fuzzy). The principal modalities of generalized constraints are summarized in the following.

2.1. Principal modalities of generalized constraints.

(a) Possibilistic (*r*=blank)

$$X \text{ is } R$$

with *R* playing the role of the possibility distribution of *X*. For example:

$$X \text{ is } [a, b]$$

means that $[a, b]$ is the set of possible values of *X*. Another example:

$$X \text{ is small.}$$

In this case, the fuzzy set labeled small is the possibility distribution of *X*. If μ_{small} is the membership function of small, then the semantics of "*X* is small" is defined by

$$\text{Poss}\{X=u\} = \mu_{small}(u)$$

where *u* is a generic value of *X*.

(b) Probabilistic (*r*=*p*)

$$X \text{ isp } R,$$

with *R* playing the role of the probability distribution of *X*. For example.

$$X \text{ isp } N(m, \sigma^2)$$

means that *X* is a normally distributed random variable with mean *m* and variance σ^2.

If *X* is a random variable which takes values in a finite set $\{u_1, \ldots, u_n\}$ with respective probabilities $p_1, \ldots, p_n$, then *X* may be expressed symbolically as

$$X \text{ isp } (p_1 \backslash u_1 + \ldots + p_n \backslash u_n),$$

with the semantics

$Prob(X=u_i)=p_i, \quad i=1, \ldots, n.$

What is important to note is that in GTU a probabilistic constraint is viewed as an instance of a generalized constraint.

When X is a generalized constraint, the expression

X isp R

is interpreted as a probability qualification of X, with R being the probability of X, Zadeh [44]. For example.

(X is small) isp likely,

where small is a fuzzy subset of the real line, means that the probability of the fuzzy event {X is small} is likely. More specifically, if X takes values in the internet [a, b] and g is the probability density function of X, then the probability of the fuzzy even "X is small" may be expressed as (Zadeh [40])

$$\text{Prob}(X \text{ is small}) = \int_a^b \mu_{small}(u)g(u)du$$

Hence

$$ts(g) = \mu_{likely}(\int_b^a g(u)\mu_{small}(u)).$$

This expression for the test-score function defines the semantics of probability qualification of a possibilistic constraint.

<u>Veristic</u> ($r=v$)

X isv R,

where R plays the role of a verity (truth) distribution of X. In particular, if X takes values in a finite set $\{u_1, \ldots, u_n\}$ with respective verity (truth) values $t_1, \ldots, t_n$, then X may be expressed as

X isv $(t_1|u_1 + \ldots + t_n|u_n)$,

meaning that $Ver(X=u_i)=t_i, \quad i=1, \ldots, n.$

For example, if Robert is half German, quarter French and quarter Italian, then

Ethnicity(Robert) isv 0.5|German+0.25|French+0.25|Italian.

When X is a generalized constraint, the expression

X isv R

is interpreted as verity (truth) qualification of *X*. For example,

(X is small) isv very.true,

should be interpreted as "It is very true that *X* is small." The semantics of truth qualification is defined in (Zadeh [40])

$$\text{Ver}(X \text{ is } R) \text{ is } t \longrightarrow X \text{ is } \mu_R^{-1}(t),$$

where μ_R^{-1} is inverse of the membership function of R_i and t is a fuzzy truth value which is a subset of [0, 1], Fig. 3.

Note: There are two classes of fuzzy sets: (a) possibilistic, and (b) veristic. In the case of a possibilistic fuzzy set, the grade of membership is the degree of possibility. In the case of a veristic fuzzy set, the grade of membership is the degree of verity (truth). Unless stated to the contrary, a fuzzy set is assumed to be possibilistic.

Usuality ($r=u$)

X isu R.

The usuality constraint presupposes that *X* is a random variable, and that the probability of the event {*X* isu *R*} is usually, where usually plays the role of a fuzzy probability which is a fuzzy number (Kaufman and Gupta [19]. For example.

X isu small

means that "usually *X* is small" or, equivalently,

Prob{*X* is small} is usually.

In this expression, small may be interpreted as the usual value of X. The concept of a usual value has the potential of playing a significant role in decision analysis, since it is more informative than the concept of an expected value.

Random-set constraint ($r=vs$)

In

X isrs R,

X is a fuzzy-set-valued random variable and *R* is a fuzzy random set

Fuzzy-graph constraint (*r=fq*)

In

X isfg R,

X is a function, f, and R is a fuzzy graph (Zadeh [51]) which constrains f (Fig. 4). A fuzzy graph is a disjunction of Cartesian granules expressed as

$$R=A_1\times B_1+\ldots+A_n\times B_n,$$

where the A_i and B_1, i=1, …, n, are fuzzy subsets of the real line, and × is the Cartesian product. A fuzzy graph is frequently described as a collection of fuzzy if-then rules (Zadeh[52], Pedrycz and Gomide [29], Bardossy and Duckstein [2]).

R: if X is A_1 then Y is B_1, i=1, …, n

The concept of a fuzzy-graph constraint plays an important role in applications of fuzzy logic [2], [15], [18].

Bimodal (*r=bm*)

In the bimodal constraint,

X isbm R,

R is a bimodal distribution of the form

R: Σ_i $P_i\backslash A_i$, i=1, …, n.

which means that Prob(X is A_i) is P_i. (Zadeh [55])

Example:

R: low\small+high\medium+low\large.

There are two types of bimodal distributions. In type 1, X is a real-valued random variable; the A_i are fuzzy subsets of the real line; and the P_i are granular probabilities of the A_i (Fig. 5). Thus

Prob(X is A_i) is P_i, i=1, …, n.

In type 2, X is a fuzzy-set-valued random variable taking the values A_1, …, A_n with respective granular probabilities P_1, …, P_n. Unless stated to the contrary, a bimodal distribution is assumed to be of type 1.

The importance of bimodal distributions derives from the fact that in many realistic settings a bimodal distribution is the best approximation to our state of knowledge. An example is assessment of degree of relevance, since relevance is generally not well defined. If I am asked to assess the degree of relevance of a book on knowledge representation to summarization, my state of knowledge about the book may not be sufficient to justify an answer such as 0.7. A better approximation to my state of knowledge may be "likely to be high." Such an answer is an instance of a bimodal distribution.

What is the expected value of a bimodal distribution? This question is considered in Section 5.

Group $(r=g)$

In

X isg R,

X is a group variable, $G[A]$, and R is a group constraint on $G[A]$. More specifically, if X is a group variable of the form

$G[A]$: $\text{Name}_1/A_i + \ldots + \text{Name}_n/A_n$

or

$G[A]$: $\Sigma_i \text{Name}_1/A_i$, for short, $i=1, \ldots, n$,

then R is a constraint on the A_i. To illustrate, if we have a group of n Swedes, with Name_i being the name of ith Swede, and A_i being the height of Name_i, then the proposition "most Swedes are tall," is a constraint on the A_i which may be expressed as (Zadeh [25])

$$\frac{1}{n}\sum Count(tall.Swedes) \text{ is most}$$

or, more explicitly,

$$\frac{1}{n}(\mu_{tall}(A_1)+\ldots+\mu_{tall}(A_n)) \text{ is most,}$$

where most is a fuzzy quantifier which is interpreted as a fuzzy number

2.2. Operations on generalized constraints

There are many ways in which generalized constraints may be operated on.

The basic operations—expressed in symbolic form—are the following.

Conjunction

$$\frac{\begin{array}{l} X \text{ isr } R \\ Y \text{ iss } S \end{array}}{(X,Y) \text{ ist } T}$$

Example (possibilistic constraints) (Fig. 6)

$$\frac{\begin{array}{l} X \text{ is } R \\ Y \text{ is } S \end{array}}{(X,Y) \text{ is } R \times S}$$

where × is the Cartesian product.

Example (probabilistic/possibilistic)

$$\frac{\begin{array}{l} X \text{ isp } R \\ (X,Y) \text{ is } S \end{array}}{(X,Y) \text{ isrs } T}$$

In this example, if *S* is a fuzzy relation then *T* is a fuzzy random set. What is involved in this example is a conjunction of a probabilistic constraint and a possibilistic constraint. This type of probabilistic/possibilistic constraint plays a key role in the Dempster-Shafer theory of evidence, and in its extension to fuzzy sets and fuzzy probabilities (Zadeh [43]).

Example (possibilistic/probabilistic)

$$\frac{\begin{array}{l} X \text{ is } R \\ (X,Y) \text{ isp } S \end{array}}{Y/X \text{ isp } T}$$

This example, which is a dual of the proceeding example, is an instance of conditioning.

Projection *(possibilistic)* (Fig. 7)

$$\frac{(X,Y) \text{ is } R}{X \text{ is } S} ,$$

where *X* takes values in *U*={*u*}; *Y* takes values in *V*={*v*}; and the projection

$$S = \text{Proj}_x R,$$

is defined as

$$\mu_S(v) = \mu_{Proj_X R}(v) = max_v \, \mu_r(u,v),$$

where μ_R and μ_S are the membership functions of *R* and *S*, respectively.

Projection (probabilistic)

$$\frac{(X,Y)\ \text{isp}\ R}{X\ \text{isp}\ S}\ ,$$

where *X* and *Y* are real-valued random variables, and *R* and *S* are the probability distributions of (*X*,*Y*) and *X*, respectively. The probability density function of *S*, p_S, is related to that of *R*, p_R, by the familiar equation

$$p_S(u) = \int p_R(u,v)dv$$

with the integral taken over the real line.

Propagation

$$\frac{f(X)\ \text{isr}\ R}{g(X)\ \text{iss}\ S}\ ,$$

where *f* and *g* are functions or functionals.

Example (possibilistic constraints) (Fig. 8)

$$\frac{f(X)\ \text{is}\ R}{g(X)\ \text{is}\ S}$$

where *R* and *S* are fuzzy sets. In terms of the membership function of *R*, the membership function of *S* is given by the solution of the variational problem

$$\mu_S(v) = sup_u(\mu_R(f(u))$$

subject to

$$v = g(u)\ .$$

Note: The constraint propagation rule described in this example is the well-known extension principle of fuzzy logic, Zadeh [39, 42]. Basically, this principle provides a way of computing the possibilistic constraint on *g*(*X*) given a possibilistic constraint on *f*(*X*).

2.3. Primary constraints, composite constraints and the Generalized Constraint Language (GCL)

Among the principal generalized constraints there are three that play the role of primary generalized constraints. They are:

Possibilistic constraint: X is R

Probabilistic constraint: X isp R

and

Veristic constraint: X isv R

A generalized constraint, GC, is composite if it can be generated from other generalized constraints through conjunction, and/or projection and/or constraint propagation and/or qualification and/or possibly other operations. For example, a random-set constraint may be viewed as a conjunction of a probabilistic constraint and either a possibilistic or veristic constraint. The Dempster-Shafer theory of evidence is, in effect, a theory of possibilistic random-set constraints. The derivation graph of a composite constraint defines how it can be derived from primary constraints.

The three primary constraints—possibilistic, probabilistic and veristic—are closely related to a concept which has a position of centrality in human cognition—the concept of partiality. In the sense used here, partial means: a matter of degree or, more or less equivalently, fuzzy. In this sense, almost all human concepts are partial (fuzzy). Familiar examples of fuzzy concepts are: knowledge, understanding, friendship, love, beauty, intelligence, belief, causality, relevance, honesty, mountain and, most important, truth, likelihood and possibility. Is a specified concept, *C*, fuzzy? A simple test is: If *C* can be hedged, then it is fuzzy. For example, in the case of relevance, we can say: very relevant, quite relevant, slightly relevant, etc. Consequently, relevance is a fuzzy concept.

The three primary constraints may be likened to the three primary colors: red, blue and green. In terms of this analogy, existing themes of uncertainty may be viewed as theories of different mixtures of primary constraints. For example, the Dempster-Shafer theory of evidence is a theory of a mixture of probabilistic and possibilistic constraints. The Generalized Theory of Uncertainty embraces all possible mixtures, and in this sense the conceptual structure of GTU accommodates most, and perhaps all, of the existing theories of uncertainty.

2.4. The Generalized Constraint Language

A concept which plays an important role in GTU is that of Generalized Constraint Language (GCL). Informally, GCL is the set of all generalized constraints together with the rules governing syntax, semantics and generation. Simple examples of elements of GCL are:

$((X,Y) \text{ isp } A) \wedge (X \text{ is } B)$

$(X \text{ isp } A) \wedge ((X,Y) \text{ isv } B)$

$\text{Proj}_Y((X \text{ is } A) \wedge (X,Y) \text{ isp } B)$

where $\wedge$ is conjunction.

A very simple example of a semantic rule is:

$$(X \text{ is } A) \wedge (Y \text{ is } B) \longrightarrow \text{Poss}(X{=}u, Y{=}v) = \mu_A(u) \wedge \mu_B(v),$$

where u and v are generic values of X, Y, and μ_A and μ_B are the membership functions of A and B, respectively.

In principle, GCL is an infinite set. However, in most applications only a small subset of GCL is likely to be needed.

3. The Concept of Precisiation and PNL

How can precise meaning be assigned to a proposition, p, drawn from a natural language?

The problem is that natural languages are intrinsically imprecise. Imprecision of natural languages is a consequence of the fact that (a) a natural language is, basically, a system for describing perceptions; and (b) perceptions are intrinsically imprecise as a consequence of (a) the bounded ability of sensory organs, and ultimately the brain, to resolve detail and store information; and (b) incompleteness of information.

Given these facts, how can we precisiate the meaning p?

A key idea which underlies the concept of Precisiated Natural Language (PNL), Zadeh [55] is to represent the meaning of p as a generalized constraint, Fig. 9. In symbols.

$$p \longrightarrow X \text{ isr } R.$$

This idea is consistent with the fundamental premise of GTU, namely, that information is representable as a generalized constraint. The basis for the consistency is that a proposition, viewed as an answer to a question, is a carrier of information. In this sense, the premise “Information is representable as a generalized constraint,” is equivalent to the premise “A proposition is representable as a generalized constraint.” A forerunner of PNL is PRUF (Zadeh[48]).

Given that the Generalized Constraint Language, GCL, is the set of all generalized constraints, representing *p* as a generalized constraint is equivalent to translating *p* into an element, *p**, of GCL. Thus, precisiation of a natural language, NL, may be viewed as translation of NL into GCL. Equivalently, translation of *p* into GCL may be viewed as explicitation of *X*, *R* and *r*, Fig. 10.

A proposition, *p*, is precisiable if it is translatable into GCL. Not every proposition in NL is precisiable. But since GCL includes every possible constraint, it is more expressive in relation to NL than any existing synthetic language, among them the languages associated with first order logic, modal logic, Prolog and LISP.

Translation of *p* into GCL is made more transparent though annotation. To illustrate,

(a) *p*: Monika is young $\longrightarrow$ *X*/Age(Monika) is *R*/young

(b) *p*: It is likely that Monika is young $\longrightarrow$ Prob(*X*/Age(Monika) is *R*/young) is *S*/likely

<u>Note</u>: Example (b) is an instance of probability qualification.

More concretely, let *g*(*u*) be the probability density function of the random variable, Age(Monika). Then, with reference to our earlier discussion of probability qualification, we have

Prob(Age(Monika) is young) is likely $\longrightarrow$

$$\int_0^{100} g(u)\mu_{young}(u)du \quad \text{is} \quad \text{likely}$$

or, in annotated form,

$$\text{GC(g)}= X \,/ \int_0^{100} g(u)\mu_{young}(u)du \;\; \text{is } R\text{/likely.}$$

The test-score of this constraint on *g* is given by

$$ts(g)=\mu_{likely}\left(\int_0^{100} g(u)\mu_{young}(u)du\right).$$

(c) *p*: Most Swedes are tall.

Following (b), let *h*(*u*) be the count density function of Swedes, meaning that

h(*u*)*du*= fraction of Swedes whose height lies in the interval [*u*, *u*+*du*].

Assume that height of Swedes lies in the interval [a, b]. Then, fraction of tall Swedes: $\int_a^b h(u)\mu_{tall}(u)du$ is most.

Interpreting this relation as a generalized constraint on *h*, the test-score may be expressed as

$$ts(h)=\mu_{most}(\int_0^h h(u)\mu_{tall}(u)du).$$

In summary, precisiation of “Most Swedes are tall” may be expressed as the generalized constraint.

$$\text{Most Swedes are tall} \longrightarrow GC(h)=\mu_{likely}(\int_a^b h(u)\mu_{tall}(u)du).$$

An important application of the concept of precisiation relates to precisiation of propositions of the form “*X* is approximately *a*,” where *a* is a real number. How can “approximately *a*”, or **a* for short, be precisiated? In other words, how can the uncertainty associated with the value of *X* which is described as **a*, be defined precisely?

There is a hierarchy of ways in which this can be done. The simplest is to define **a* as *a*. This mode of precisiation will be referred to as singular precisiation, or *s*-precisiation, for short (Fig. 11) *s*-precisiation is employed very widely, especially in probabilistic computations, in which an imprecise probability, **a*, is computed with as if it were an exact number, *a*.

The other ways (Fig. 12) will be referred to as granular precisiation, or *g*-precisiation, for short. In *g*-precisiation, **a* is treated as a granule. What we see is that various modes of precisiating **a* are instances of the generalized constraint.

The concept of precisiation has an inverse—the concept of imprecisiation, which involves replacing *a* with **a*, with the understanding that *a is not unique.

A basic problem which relates to imprecisiation is the following. Assume for simplicity that we have two linear equations involving real-valued coefficients and real-valued variables:

$$a_{11}X+a_{12}Y=b_1$$
$$a_{21}X+a_{22}Y=b_2.$$

Solutions of these equations read,

$$X=\frac{a_{22}b_1-a_{12}b_2}{a_{11}a_{22}-a_{12}a_{21}}$$

$$Y=\frac{a_{11}b_2-a_{21}b_1}{a_{11}a_{22}-a_{12}a_{21}}.$$

Now suppose that we imprecisiate the coefficients, replacing, a_{ij} with $*a_{ij}$, i, j= 1, 2, and replacing b_i with $*b_i$, i= 1,2. How can we solve these equations when imprecisiated coefficients are defined as generalized constraints?

There is no general answer to this question. Assuming that all coefficients are defined in the same way, the method of solution will depend on the modality of the constraint. For example, if the coefficients are interval-valued, the problem falls within the province of interval analysis (Moore [24]). If the coefficients are fuzzy-interval-valued, the problem falls within the province of the theory of relational equations (Di Nola et al [6, 7], Mares [23]). And if the coefficients are real-valued random variables, we are dealing with the problem of solution of stochastic equations. In general, solution of a system of equation with imprecisiated coefficients may present complex problems.

One complication is the following. If (a) we solve the original equations, as we have done above; (b) imprecisiate the coefficients in the solution; and (c) employ the extension principle to complete X and Y, will we obtain solutions of imprecisiated equations? The answer, in general, is: No.

Nevertheless, when we are faced with a problem which we do not know how to solve correctly, we proceed as if the answer is: Yes. This common practice may be described as Precisiation/Imprecisiation Principle which is defined in the following.

3.1. Precisiation/Imprecisiation Principle (P/I Principle)

Informally, let f be a function or a functional. $Y=f(X)$, where X and Y are assumed to be imprecise, $Pr(X)$ and $Pr(Y)$ are precisiations of X and Y, and $*Pr(X)$ and $*Pr(Y)$ are imprecisiations of $Pr(X)$ and $Pr(Y)$, respectively. In symbolic form, the P/I Principle may be expressed as

$$f(X)*=*f(Pr(X))$$

where *= denotes “approximately equal,” and $*f$ is imprecisiation of f. In words, to compute $f(X)$ when X is imprecise, (a) precisiate X, (b) compute $f(Pr(X))$; and (c) imprecisiation $f(Pr(X))$. Then, usually, $*f(Pr(X))$ will be approximately equal to $f(X)$. An underlying assumption is that approximation, are commensurate in the sense that the closer Pr(X) is to X, the closer f(Pr(X)) is to f(X). This assumption is related to the concept of gradual rules of Dubois and Prade [9].

As an illustration, suppose that X is a real-valued function; f is the operation of differentiation, and $*X$ is the fuzzy graph of X. Then, using the using the P/I Principle, $*f(X)$ will have the form shown in Fig.13. It should be underscored that imprecisiation is an imprecise concept.

Use of the P/I Principle underlies many computations in science, engineering, economics and other fields. In particular, as was alluded to earlier, this applies to many computation in probability theory which involve imprecise probabilities. It should be emphasized that the P/I Principle is neither normative (prescriptive) nor precise; it merely describes imprecisely what is common practice—without suggesting that common practice is correct.

4. Precisiation of Propositions

In the preceding section, we focused our attention on precisiation of propositions of the special form "*X* is **a*." In the following, we shall consider precisiation in a more general setting. In this setting, the concept of precisiation in PNL opens the door to a wide-ranging enlargement of the role of natural languages in scientific theories, especially in fields such as economics, law and decision analysis. Our discussion will be brief; details may be found in Zadeh [56]

Precisiation of propositions—and the related issues of precisiation of questions, commands and concepts—fall within the province of PNL (Precisiated Natural Language). As was stated earlier, the point of departure in PNL is representation of the meaning of a proposition, *p*, as a generalized constraint.

$$p \longrightarrow X \text{ isr } R.$$

To illustrate precisiation of propositions and questions, it will be convenient to consider the examples which were discussed earlier in Section 1.

The Robert example

p: Usually Robert returns from work at about 6 pm.
q: What is the probability that Robert is home at about 6:15 pm?

Precisiation of *p* may be expressed as

p: Prob(Time(Return(Robert)) is *6:00 pm) is usually

where "usually" is a fuzzy probability

Assuming that Robert stays home after returning from work, precisiation of *q* may be expressed as

q: Prob(Time(Return(Robert)) is $\leq \circ$ 6:15 pm) is *A*?

where $\circ$ is the operation of composition, and *A* is a fuzzy probability

The balls-in-box problem

p_1: A box contains about 20 black and white balls
p_2: Most are black

p_3: There are several times as many black balls as white balls
q_1: What is the number of white balls?
q_2: What is the probability that a ball drawn at random is white?

Let X be the number of black balls and let Y be the number of white balls. Then, in precisiated form, the statement of the problem may be expressed as:

$$\left.\begin{array}{l} p_1: (X+Y) \text{ is } {}^*20 \\ p_2: X \text{ is most} \times {}^*20 \\ p_3: X \text{ is several} \times Y \end{array}\right\} \text{data}$$

$$\left.\begin{array}{l} q_1: Y \text{ is } ?A \\ q_2: \dfrac{Y}{{}^*20} \text{ is } ?B \end{array}\right\} \text{questions}$$

where Y/*20 is the granular probability that a ball drawn at random is white

Solution of these equations reduces to an application of fuzzy integer programming (Fig.14).

The tall Swedes problem

p: Most Swedes are tall.
q: What is the average height of Swedes?
q: How many Swedes are short?

As was shown earlier,

$$p: \text{Most Swedes are tall} \longrightarrow \int_a^b h(u)\mu_{tall}(u)du \text{ is most}$$

where h is the count density function.

Precisiations of q_1, and q_2 may be expressed as

$$q_1: \int uh(u)du \text{ is } ?A$$

where A is a fuzzy number which represents the average height of Swedes, and

$$q_2: \int h(u)\mu_{short}(u)du \text{ is } ?B$$

where μ_{short} is the membership function of short, and B is the fraction of short Swedes.

The partial existence problem

X is a real number. I am uncertain about the value of X. What I know about X is:

p_1: X is much larger than approximately a
p_2: X is much smaller than approximately b

where a and b are real numbers, with $a < b$.
What is the value of X?

In this case, precisiations of data may be expressed as

p_1: X is much.larger $\circ$ $*a$
p_2: X is much smaller $\circ$ $*b$

where $\circ$ is the operation of compostion. Precisiation of the question is:

q: X is $?A$

where A is a fuzzy number. The solution is immediate:

X is much.larger $\circ$ $*a \wedge$ much.smaller $\circ$ $*b$

when $\wedge$ is min or a t-norm. In this instance, depending on a and b, X may exist to a degree.

These examples point to an important aspect of precisiation. Specifically, to precisiate p we have to precisiate or, equivalently, calibrate its lexical constituents. For example, in the case of “Most Swedes are tall,” we have to calibrate “most” and “tall.” (Fig. 7) Likewise, in the case of the Robert example, we have to calibrate “about 6:00 pm,” “about 6:15 pm” and “usually.” In effect, we are composing the meaning of p from the meaning of its constituents. This process is in the spirit of Frege’s principle of compositionality, Zadeh [47], Montague grammar [28] and the semantics of programming languages.

An important aspect of precisiation which will not be discussed here relates to precisiation of concepts. It is a deep-seated tradition in science to base definition of concepts on bivalent logic. In probability theory, for example, independence of events is a bivalent concept. But, in reality, independence is a matter of degree, i.e., is a fuzzy concept. PNL, used as a definition language, makes it possible, more realistically, to define independence and other bivalent concepts in probability theory as fuzzy concepts. For this purpose, when PNL is used as a definition language, a concept is first defined in a natural language and then its definition is precisiated through the use of PNL.

5. <u>Reasoning Under Uncertainty</u>

Reasoning under uncertainty has many facets. The facet that is the primary focus of attention in GTU is reasoning with, or equivalently, deduction from, uncertain information expressed in a natural language.

Precisiation is a prelude to deduction. In this context, deduction in GTU involves, for the most part, computation with precisiations of propositions drawn from a natural language. A concept which plays a key role in deduction is that of a protoform—abbreviation of “prototypical form” (Zadeh [56]).

<u>The concept of a protoform</u>

Informally, a protoform of an object is its abstracted summary. More specifically, a protoform is a symbolic expression which defines the deep semantic structure of an object such as a proposition, question, command, concept, scenario, case or a system of such objects. In the following, our attention which will be focused on protoforms of propositions, with PF(*p*) denoting a protoform of *p*. (Fig.15). Abstraction has levels, just as summarization does. For this reason, an object may have a multiplicity of protoforms (Fig.16). Conversely, many objects may have the same protoform. Such objects are said to be protoform-equivalent, or PF-equivalent, for short. The set of protoforms of all precisiable propositions in NL, together with rules which govern propagation of generalized constraints, constitute what is called the Protoform Language (PFL).

Examples:

instantiation

- Monika is young ⟶ Age(Monika) is young ⟶ $A(B)$ is C

abstraction

- Monika is much younger than Pat ⟶ $(A(B), A(C))$ is R

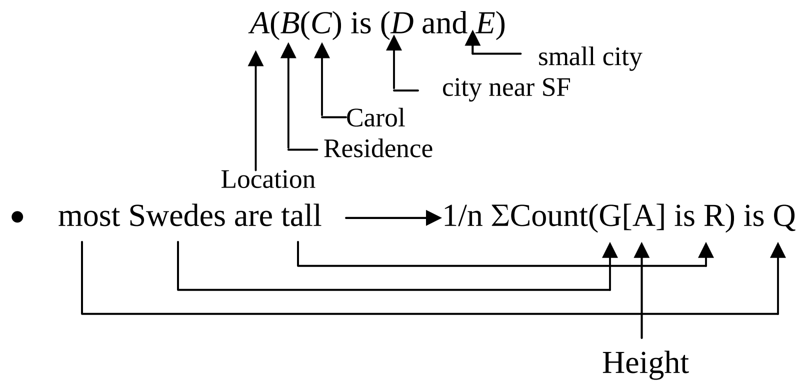
Age Monika Age Pat much younger

- distance between New York and Boston is about 200 mi ⟶ $A(B, C)$ is R
- usually Robert returns from work at about 6pm ⟶

Prob{A is B} is C

usually
about 6 pm
Time(Robert.returns.from.work)

- Carol lives in a small city near San Francisco ⟶



Protoformal deduction

The rules of deduction in GTU are, basically, the rules which govern constraint propagation. In GTU, such rules reside in the Deduction Database (DDB), Fig.18. The Deduction Database comprises a collection of agent-controlled modules and submodules, each of which contains rules drawn from various fields and various modalities of generalized constraints. A typical rule has a symbolic part, which is expressed in terms of protoforms; and a computational part which defines the computation that has to be carried out to arrive at a conclusion. In what follows, we describe briefly some of the basic rules, and list a number of other rules without describing their computational parts. The motivation for doing so is to point to the necessity of developing a set of rules which is much more complete than the few rules which are used as examples in this section.

(a) Computational rule of inference (Zadeh [18])

Symbolic part	Computational part
X is A	
(X, Y) is B	$\mu_C(v) = max_u(\mu_A(u) \wedge \mu_B(u,v))$
Y is C	

A, B and C are fuzzy sets with respective membership functions $\mu_A \mu_B \mu_C$ $\wedge$ is min or t-norm (Fig. 19).

(b) Intersection / product syllogism (Zadeh [46])

Symbolic part	Computational part
Q_1 A's are B's	
Q_2 $(A\&B)$'s are C's	$Q_3 = Q_1 * Q_2$
Q_3 A's are $(B\&C)$'s	

Q_1 and Q_2 are fuzzy quantifiers; *A*, *B*, *C* are fuzzy sets; $*$ is product in fuzzy arithmetic. [14]

(c) Basic extension principle (Zadeh [39])

Symbolic part	Computational part
X is *A*	$\mu_B(v) = sup_u(\mu_A(u))$
f(*X*) is *B*	subject to
	$v = g(u)$

g is a given function or functional; *A* and *B* are fuzzy sets. (Fig.20)

Extension principle (Zadeh [55])

This is the principal rule governing possibilistic constraint propagation (Fig.8)

Symbolic part	Computational part
f(*X*) is *A*	$\mu_B(v) = sup_u(\mu_B(f(u)))$
g(*X*) is *B*	subject to
	$v = g(u)$

Note: The extension principle is a primary deduction rule in the sense that many other deduction rules are derivable from the extension principle. An example is the following rule.

(d) Basic probability rule

Symbolic part	Computational part
Prob(*X* is *A*) is *B*	$\mu_D(v) = sup_g(\mu_B(\int_U \mu_A(u)g(u)du))$
Prob(*X* is *C*) is *D*	subject to
	$v = \int_U \mu_C(u)r(u)du$
	$\int_U r(u)du = 1.$

X is a real-valued random variable; A, B, C and D are fuzzy sets: r is the probability density of X; and $U=\{u\}$. To derive this rule, we note that

$$\text{Prob}(X \text{ is } A) \text{ is } B \longrightarrow \int_U r(u)\mu_A(u)du \text{ is } B$$
$$\text{Prob}(X \text{ is } C) \text{ is } D \longrightarrow \int_U r(u)\mu_C(u)du \text{ is } D$$

which are generalized constraints of the form

$f(r)$ is B
$g(r)$ is D.

Applying the extension principle to these expressions, we obtain the expression for D which appears in the basic probability rule

(e) Bimodal interpolation rule

The bimodal interpolation rule is a rule which resides in the Probability module of DDB. With reference to Fig. 21, the symbolic and computational parts of this rule are:

Symbolic

$$\frac{\text{Prob}(X \text{ is } A_i) \text{ is } P_i}{\text{Prob}(X \text{ is } A) \text{ is } Q} \quad , i=1, \ldots, n$$

Computational

$$\mu_Q(v) = sup_r(\mu_{P_1}(\int_U \mu_{A_1}(u)r(u)du)) \wedge \ldots \wedge \mu_{P_n}(\int_U \mu_{A_n}(u)r(u)du))$$

subject to

$$v = \int_U \mu_A(u)r(u)du$$

$$\int_U r(u)du = 1$$

In this rule, X is a real-valued random variable; r is the probability density of X; and U is the domain of X.

Note: The probability rule is a special case of the bimodal interpolation rule.

What is the expected value, $E(X)$, of a bimodal distribution? The answer follows through application of the extension principle:

$$\mu_{E(X)}(v) = sup_r(\mu_{P_1}(\int_U \mu_{A_1}(u)r(u)du)) \wedge \ldots \wedge \mu_{P_n}(\int_U \mu_{A_n}(u)r(u)du))$$

subject to

$$v = \int_U ur(u)du$$

$$\int_U r(u)du = 1$$

<u>Note</u>: $E(X)$ is a fuzzy subset of U.

(f) <u>Fuzzy-graph interpolation rule</u>

This rule is the most widely used rule in applications of fuzzy logic (Zadeh [51]). We have a function, $Y = f(X)$, which is represented as a fuzzy graph (Fig.22). The question is: What is the value of Y when X is A? The A_i, B_i and A are fuzzy sets.

Symbolic part

X is A
$Y = f(X)$
$f(X)$ isfg $\sum_i A_i \times B_i$
Y is C

Computational part

$C = \sum_i m_i \wedge B_i$,

where m_i is the degree to which A matches A_i

$$m_i = sup_u(\mu_A(u) \wedge \mu_{A_i}(u)) \quad , \qquad i=1, \ldots, n.$$

When A is a singleton, this rule reduces to

$X = a$
$Y = f(X)$
$f(X)$ isfg $\sum_i A_i \times B_i$, $i=1, \ldots, n.$
$Y = \sum_i \mu_{A_i}(a) \wedge B$;

In this form, the fuzzy-graph interpolation rule coincides with the Mamdani rule—a rule which is widely used in control and related applications. (Mamdani and Assilian [22]) Fig. 23.

In the foregoing, we have summarized some of the basic rules in DDB which govern generalized constraint propagation. Many more rules will have to be developed and added to DDB. A few examples of such rules are the following.

(a) Probabilistic extension principle

$$\frac{f(X) \text{ isp } A}{g(X) \text{ isr } ?B}$$

(b) Usuality-qualified extension principle

$$\frac{f(X) \text{ isu } A}{g(X) \text{ isr } ?B}$$

(c) Usuality-qualified fuzzy-graph interpolation rule

$$\frac{\begin{array}{l} X \text{ is } A \\ Y = f(X) \\ f(X) \text{ isfg } \sum_i \text{ if } X \text{ is } A_i \text{ then } Y \text{ isu } B_i \end{array}}{Y \text{ isr } ?\, B}$$

(d) Bimodal extension principle

$$\frac{\begin{array}{l} X \text{ isbm } \sum_i P_i \backslash A_i \\ Y = f(X) \end{array}}{Y \text{ isr } ?\, B}$$

(e) Bimodal, binary extension principle

$$\frac{\begin{array}{l} X \text{ isr } R \\ Y \text{ iss } S \\ Z=f(X,Y) \end{array}}{Z \text{ ist } T}$$

In the instance, bimodality means that X and Y have different modalities, and binary means that f is a function of two variables. An interesting special case is one in which X is R and Y isp S.

The deduction rules which were briefly described in the foregoing are intended to serve as examples. How can these rules be applied to reasoning under uncertainty? To illustrate, it will be convenient to return to the examples given in section 1.

The Robert example

p: Usually Robert returns from work at about 6:00 pm. What is the probability that Robert is home at 6:15 pm?

First, we find the protoforms of the data and the query.

Usually Robert returns from work at about 6:00 pm ⟶

⟶ Prob(Time(Return(Robert)) is *6:00 pm) is usually

which in annotated form reads

⟶ Prob(*X* /Time(Return(Robert)) is *A*/*6:00pm) is *B*/usually

Likewise, for the query, we have

Prob(Time(Return(Robert)) is ≤ ∘ *6:15pm) is ? *D*

which in annotated form reads

⟶ Prob(*X*/Time(Return(Robert)) is *C*/ ≤ ∘ *6:15pm) is *D*/usually

Searching the Deduction Database, we find that the basic probability rule matches the protoforms of the data and the query

$$\frac{\text{Prob } (X \text{ is } A) \text{ is } B}{\text{Prob } (X \text{ is } C) \text{ is } D}$$

where

$$\mu_D(v) = sup_g(\mu_B(\int_U \mu_A(u)g(u)du))$$

subject to

$$v = \int_U \mu_C(u)g(u)du$$

$$\int_U g(u)du = 1$$

Instantiating *A*, *B*, *C* and *D*, we obtain the answer to the query:

Probability that Robert is home at about 6:15pm is *D*,

where

$$\mu_D(v) = sup_g(\mu_{usually}(\int_U \mu_{*6:00\,pm}(u)g(u)du))$$

subject to

$$v = \int_U \mu_{\leq \circ * 6:15\,pm}(u)g(u)du$$

and

$$\int_U g(u)du = 1$$

<u>The tall Swedes problem</u>

We start with the data

p: Most Swedes are tall.

Assumes that the queries are:

q_1: How many Swedes are not tall

q_2: How many are short

q_3: What is the average height of Swedes

In our earlier discussion of this example, we found that *p* translates into a generalized constraint on the count density function, *h*.

Thus

$$p \longrightarrow \int_a^b h(u)\mu_{tall}(u)du \text{ is most}$$

Precisiations of q_1, q_2 and q_3 may be expressed as

$$q_1: \longrightarrow \int_a^b h(u)\mu_{not.tall}(u)du$$

$$q_2: \longrightarrow \int_a^b h(u)\mu_{short}(u)du$$

$$q_3: \longrightarrow \int_a^b \mu h(u)du\,.$$

Considering q_1, we note that

$$\mu_{not.tall}(u) = 1 - \mu_{tall}(u)\,.$$

Consequently

$$q_1 \longrightarrow 1 - \int_a^b h(u)\mu_{tall}(u)du$$

which may be rewritten as

$$q_2 \longrightarrow \text{1-most}$$

where 1-most plays the role of the antonym of most (Fig.23).

Considering q_2, we have to compute

$$A: \int_a^b h(u)\mu_{short}(u)du$$

given that $\int_a^b h(u)\mu_{tall}(u)du$ is most

Applying the extension principle, we arrive at the desired answer to the query:

$$\mu_A(v) = sup(\mu_{most}(\int_a^b h(u)\mu_{tall}(u)du))$$

subject to

$$v = \int_a^b h(u)\mu_{short}(u)du$$

and

$$\int_a^b h(u)du = 1$$

Likewise, for q_3 we have as the answer

$$\mu_A(v) = sup_u(\mu_{most}(\int_a^b h(u)\mu_{tall}(u)du))$$

subject to

$$v = \int_a^b uh(u)du$$

and

$$\int_a^b h(u)du = 1.$$

As an illustration of application of protoformal deduction to an instance of this example, consider

p: Most Swedes are tall
q: How many Swedes are short?

We start with the protoforms of p and q (see earlier example):

Most Swedes are tall ⟶ $1/n\ \Sigma\text{Count}(G[A \text{ is } R]) \text{ is } Q$
?T Swedes are short ⟶ $1/n\ \Sigma\text{Count}(G[A \text{ is } S]) \text{ is } T$
where

$G[A]=\Sigma_i \text{Name}_i/A_i, \quad i=1, \ldots, n.$

An applicable deduction rule in symbolic form is:

$$\frac{1/n\ \Sigma\text{Count}(G[A \text{ is } R]) \text{ is } Q}{1/n\ \Sigma\text{Count}(G[A \text{ is } S]) \text{ is } T}$$

The computational part of the rule is expressed as

$$\frac{1/n\ \Sigma_i\mu_R(A_i) \text{ is } Q}{1/n\ \Sigma_i\mu_S(A_i) \text{ is } T}$$

where

$$\mu_T(v) = sup_{A_i,\ldots,A_n}\ \mu_Q(\Sigma_i\mu_R(A_i))$$

subject to

$$v=\Sigma_i\mu_S(A_i)$$

What we see is that computation of the answer to the query, q, reduces to the solution of a variational problem, as it does in the earlier discussion of this example in which protoformal deduction was not employed.

<u>Vera's age problem</u>

Vera has a son who is in mid-twenties, and a daughter, who is in mid-thirties. What is Vera's age?

In dealing with this problem, we will proceed to solution directly, bypassing protoformal deduction.
Precisiations of the query and given information may be expressed as

q: What is Vera's age? ⟶ Age(Vera) is ?A
P_1: Vera has a son who is in mid-twenties ⟶ Age(Son(Vera)) is *20.
P_2: Vera has a daughter who is in mid-thirties ⟶ Age(Daughter(Vera)) is *30.

Let *X* be Vera's age when her son was born, and let *Y* be Vera's age when her daughter was born.

From World Knowledge Database, we draw the information

wk_1: Child-bearing age ranges from *16 to *42.
wk_2: Age of mother is the sum of the age of child and the age of mother when the child was born.

Combining the given information with that drawn from the World Knowledge Database, we led to an estimate of Vera's age which may be expressed as

Age(Vera) is ((*25+[*16, *42])$\wedge$(*35+[*16, *42])

The point of this example is that it underscores that, in general, computation of an estimate depends on the interpretation of "approximately *a*," when *a* is a real number. In particular, computation of Vera's age is straightforward if **a* is interpreted as a possibility distribution. It is less straightforward when *a* is interpreted as a probability distribution. And it is much less straightforward when *25, for example, is interpreted as a possibility distribution, and [*16, *42] is interpreted as a probability distribution or, more realistically, as a bimodal distribution.

The foregoing examples are merely elementary instances of reasoning through the use of generalized constraint propagation. What should be noted is that the chains of reasoning in these examples are very short. More generally, what is important to recognize is that shortness of chains of reasoning is an intrinsic characteristic of reasoning processes which take place in an environment of substantive imprecision and uncertainty. What this implies is that, in such environments, a conclusion arrived at the end of a long chain of reasoning is likely to be vacuous or of questionable validity.

Concluding Remark

Uncertainty is one of the basic facets of human cognition. Traditionally, uncertainty is dealt with through the use of tools provided by probability theory. The approach to uncertainty which is outlined in this paper suggests a much more general framework. The centerpiece of this framework is the concept of a generalized constraint, and its fundamental premise is that information may be viewed as a generalized constraint. In this perspective, probabilistic constraints are a special case—albeit an important one—of generalized constraints, and statistical information is a special case of generalized information.

Generalized constraints are large in number and variety. Computations with generalized constraints calls for a wide variety of calculi. The generalized theory of uncertainty which is outlined in this paper is merely a first step toward enhancing our understanding of the foundations of information and uncertainty.

As we enter the realm of generalized-constraint-based information and uncertainty, we find ourselves in uncharted territory. Exploration of this territory will require extensive effort and intellectual prowess. A straw in the wind is that a wide-ranging theory—the Dempster-Shafer theory of evidence—is, basically, a theory centered as just one instance of a generalized constraint—the random set constraint.

Acknowledgement

Research supported in part by ONR N00014-02-1-0294, BT Grant CT1080028046, Omron Grant, Tekes Grant and the BISC Program of UC Berkeley.

References and related papers

[1] A. Bargiela and W. Pedrycz, Granular Computing, Kluwer Academic Publishers, 2002.
[2] A. Bardossy and L. Duckstein, Fuzzy Rule-Based Modelling with Application to Geophysical, Biological and Engineering Systems, CRC Press, 1995.
[3] I. Bloch, A. Hunter, A. Appriou, A. Ayoun, S. Benferhat, P. Besnard, L. Cholvy, R. Cooke, F. Cuppens, D. Dubois, H. Fargier, M. Grabisch, R. Kruse, J. Lang, S. Moral, H. Prade, A. Saffiotti, P. Smets and C. Sossai, Fusion: General concepts and characteristics, International Journal of Intelligent Systems, 16(10): 1107-1134, 2001.
[4] B. Bouchon-Meunier, R.R. Yager and L.A. Zadeh (eds), Uncertainty in Intelligent and Information Systems, Advances in Fuzzy Systems—Applications and Theory, World Scientific, Singapore, Vol. 20, 2000.
[5] Z. Bubnicki, Analysis and Decision Making in Uncertain Systems, Springer Verlag, 2004.
[6] A. Di Nola, S. Sessa, W. Pedrycz and E. Sanchez, Fuzzy Relation Equations and Their Applications to Knowledge Engineering, Kluwer, Dordrecht, 1989.
[7] A. Di Nola, S. Sessa, W. Pedrycz and W. Pei-Zhuang, Fuzzy Relation Equation under a Class of Triangular Norms: A Survey and New Results, Fuzzy Sets for Intelligent Systems, Morgan Kaufmann Publishers, San Mateo, CA, 166-189, 1993.
[8] D. Dubois and H. Prade, Representation and combination of uncertainty with belief functions and possibility measures, Computational Intelligence 4: 244-264 1988.
[9] D. Dubois and H. Prade, Gradual inference rules in approximate reasoning, Information Sciences: an International Journal, Elsevier Science, NY, Vol. 61 (1-2): 103-122, 1992.
[10] D. Dubois, H. Fargier and H. Prade, The calculus of fuzzy restrictions as a basis for flexible constraint satisfaction, Proc. 2nd IEEE Int. Conf. on Fuzzy Systems, San Francisco, CA, 1131-1136, 1993.
[11] D. Dubois and H. Prade, Non-Standard Theories of Uncertainty in Knowledge Representation and Reasoning, KR, 634-645, 1994.
[12] D. Dubois, H. Prade (eds), John Wiley and sons, Fuzzy Information Engineering: A Guided Tour of Applications, 1996.
[13] D. Dubois, H. Fargier and H. Prade, Comparative uncertainty, belief functions and accepted beliefs, UAI, 113-120, 1998.
[14] D. Dubois and H. Prade, On the use of aggregation operations in information fusion processes, Fuzzy Sets and Systems, 142(1): 143-161, 2004.
[15] D. Filev and R.R. Yager, Essentials of Fuzzy Modeling and Control, Wiley-Interscience, 1994.
[16] J.A. Goguen, The logic of inexact concepts, Synthese, vol. 19, 325-373, 1969.
[17] M. Higashi and G.J. Klir, Measures of Uncertainty and Information Based on Possibility Distributions, Fuzzy Sets for Intelligent Systems, Morgan Kaufmann Publishers, San Mateo, CA, 217-232, 1993.

[18] M. Jamshidi, A. Titli, L.A. Zadeh and S. Boverie (eds), Applications of Fuzzy Logic—Towards High Machine Intelligence Quotient Systems, Environmental and Intelligent Manufacturing Systems Series, Prentice Hall, Upper Saddle River, NJ, Vol. 9, 1997.
[19] A. Kaufmann and M.M. Gupta, Introduction to Fuzzy Arithmetic: Theory and Applications, New York: Von Nostrand, 1985.
[20] G.J. Klir, Generalized information theory: aims, results, and open problems, Reliability Engineering and System Safety, 85(1-3), 21-38, 2004.
[21] T.Y. Lin, Granular Computing on Binary Relations-Analysis of Conflict and Chinese Wall Security Policy, In: Rough Sets and Current Trends in Computing Alpigni, Peters Skowron, Zhong (eds), Lecture Notes on Artificial Intelligence No 2475, 296-299, 2002.
[22] E.H. Mamdani and S. Assilian. An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller, Int. J. Man-Machine Studies, 7, 1-13, 1975.
[23] M. Mares, Computation Over Fuzzy Quantities, Boca Raton, FL: CRC, 1994.
[24] R.E. Moore, Interval Analyis, SIAM studies in Applied Mathematics 2, Philadelphia, PA, 1979.
[25] V. Novak, I. Perfilieva and J. Mockor, Mathematical Principles of Fuzzy Logic, Kluwer, Boston/Dordrecht, 1999.
[26] H.T. Nguyen, On Modeling of Linguistic Information Using Random Sets, Fuzzy Sets for Intelligent Systems, Morgan Kaufmann Publishers, San Mateo, CA, 242-246, 1993.
[27] H.T. Nguyen, V. Kreinovich and A. Di Nola, Which truth values in fuzzy logics are definable? International Journal of Intelligent Systems, 18(10): 1057-1064, 2003.
[28] B. Partee, Montague Grammar, New York: Academic, 1976.
[29] W. Pedrycz and F. Gomide, Introduction to Fuzzy Sets, Cambridge, MA: MIT Press, 1998.
[30] M.L. Puri and D.A. Ralescu, Fuzzy Random Variables, Fuzzy Sets for Intelligent Systems, Morgan Kaufmann Publishers, San Mateo, CA, 265-271, 1993.
[31] T.J. Ross, Fuzzy Logic with Engineering Applications, 2nd Edition, Wiley, 2004.
[32] F. Rossi, P. Codognet, Soft Constraints, Special issue on Constraints, Kluwer, vol. 8, n.1, 2003.
[33] G. Shafer, A Mathematical Theory of Evidence, Princeton, NJ: Princeton University Press, 1976.
[34] N.D. Singpurwalla and J.M. Booker, Membership functions and probability measures of fuzzy sets, Journal of the American Statistical Association, Vol. 99, No. 467, 867-889, 2004.
[35] P. Smets, Imperfect Information: Imprecision and Uncertainty, Uncertainty Management in Information Systems, 225-254, 1996.
[36] R.R. Yager, Uncertainty representation using fuzzy measures, IEEE Trans. on Systems, Man and Cybernetics, Part B. Vol. 32, 13-20, 2002.
[37] R.R. Yager, Uncertainty management for intelligence analysis, Technical Report# MII-2414 Machine Intelligence Institute, Iona College, New Rochelle, NY, 2004.
[38] J. Yen and R. Langari, Fuzzy Logic: Intelligence, Control and Information, Prentice Hall, 1st edition, 1998.

[39] L.A. Zadeh, Fuzzy sets**,** *In Control 8*, 338-353, 1965.
[40] L.A. Zadeh, Probability measures of fuzzy events, *Jour. Math. Analysis and Appl. 23*, 421-427, 1968.
[41] L.A. Zadeh, Outline of a new approach to the analysis of complex systems and decision processes, *IEEE Trans. on Systems, Man and Cybernetics SMC-3*, 28-44, 1973.
[42] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning, Part I: *Inf. Sci.8*, 199-249, 1975; Part II: *Inf. Sci. 8*, 301-357, 1975; Part III: *Inf. Sci. 9*, 43-80, 1975.
[43] L.A. Zadeh, Fuzzy sets and information granularity, *Advances in Fuzzy Set Theory and Applications*, M. Gupta, R. Ragade and R. Yager (eds.), 3-18. Amsterdam: North-Holland Publishing Co., 1979.
[44] L.A. Zadeh, A theory of approximate reasoning, *Machine Intelligence 9*, J. Hayes, D. Michie, and L.I. Mikulich (eds.), 149-194. New York: Halstead Press, 1979.
[45] L.A. Zadeh, Test-score semantics for natural languages and meaning representation via PRUF, *Empirical Semantics*, B. Rieger (ed,), 281-349. Bochum, W. Germany: Brockmeyer, 1982. Also *Technical Memorandum 246*, AI Center, SRI International, Menlo Park, CA., 1981.
[46] L.A. Zadeh, A computational approach to fuzzy quantifiers in natural languages, *Computers and Mathematics 9*, 149-184, 1983.
[47] L.A. Zadeh, A fuzzy-set-theoretic approach to the compositionality of meaning: propositions, dispositions and canonical forms, *Journal of Semantics 3*, 253-272, 1983.
[48] L.A. Zadeh, Precisiation of meaning via translation into PRUF, *Cognitive Constraints on Communication*, L. Vaina and J. Hintikka, (eds.), 373-402. Dordrecht: Reidel, 1984.
[49] L.A. Zadeh, Outline of a computational approach to meaning and knowledge representation based on a concept of a generalized assignment statement, *Proceedings of the International Seminar on Artificial Intelligence and Man-Machine Systems*, M. Thoma and A. Wyner (eds.), 198-211. Heidelberg: Springer-Verlag, 1986.
[50] L.A. Zadeh, Fuzzy control: a personal perspective, *Control Engineering*, 51-52, 1996.
[51] L.A. Zadeh, Fuzzy logic and the calculi of fuzzy rules and fuzzy graphs, *Multiple-Valued Logic 1*, 1-38, 1996.
[52] L.A. Zadeh, Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy Sets and Systems 90*, 111-127, 1997.
[53] L.A. Zadeh, Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems, *Soft Computing 2*, 23-25, 1998.
[54] L.A. Zadeh, From computing with numbers to computing with words—from manipulation of measurements to manipulation of perceptions, *IEEE Transactions on Circuits and Systems 45*, 105-119, 1999.
[55] L.A. Zadeh, Toward a perception-based theory of probabilistic reasoning with imprecise probabilities**,** *Journal of Statistical Planning and Inference*, Elsevier Science, Vol. 105, 233-264, 2002.

[56] L.A. Zadeh, Precisiated Natural Language (PNL), *AI Magazine*, Vol. 25, No. 3, 74-91, 2004.

Captions

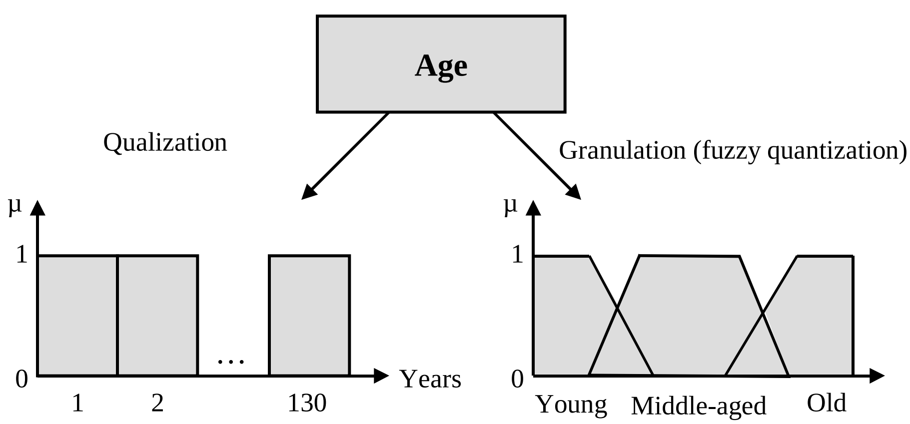
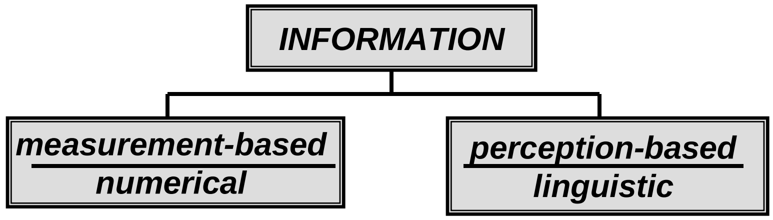


Fig. 1.



- it is 35 C°
- Over 70% of Swedes are taller than 175 cm
- probability is 0.8
-
-

- It is very warm
- Most Swedes are tall
- probability is high
- it is cloudy
- traffic is heavy
- it is hard to find parking near the campus

- ***measurement-based information may be viewed as a special case of perception-based information***
- ***perception-based information is intrinsically imprecise***

Fig. 2

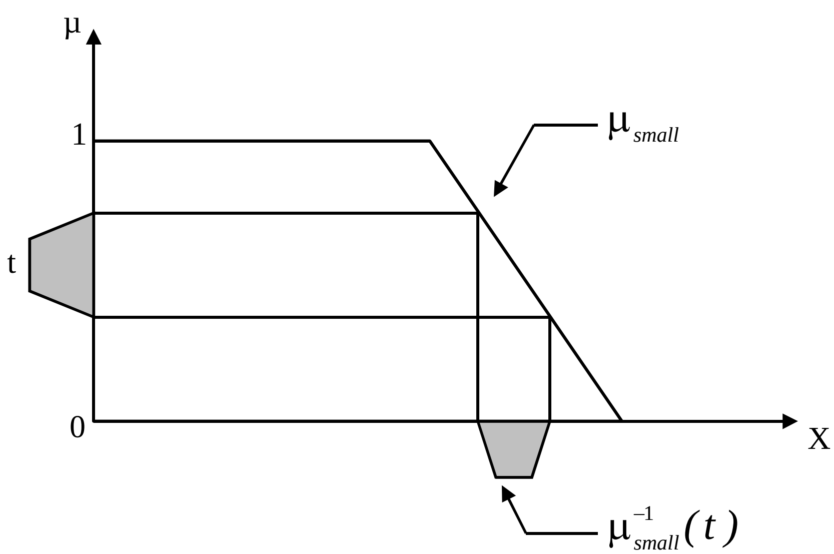
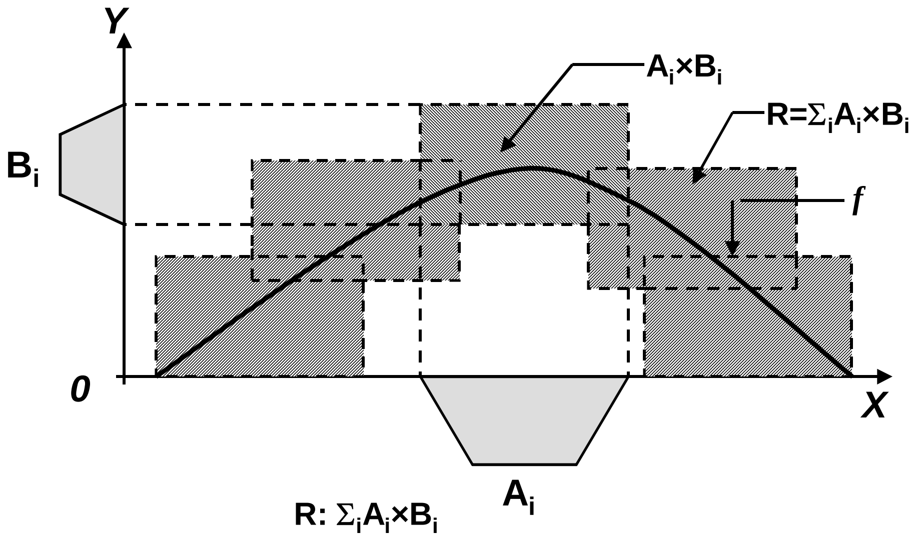
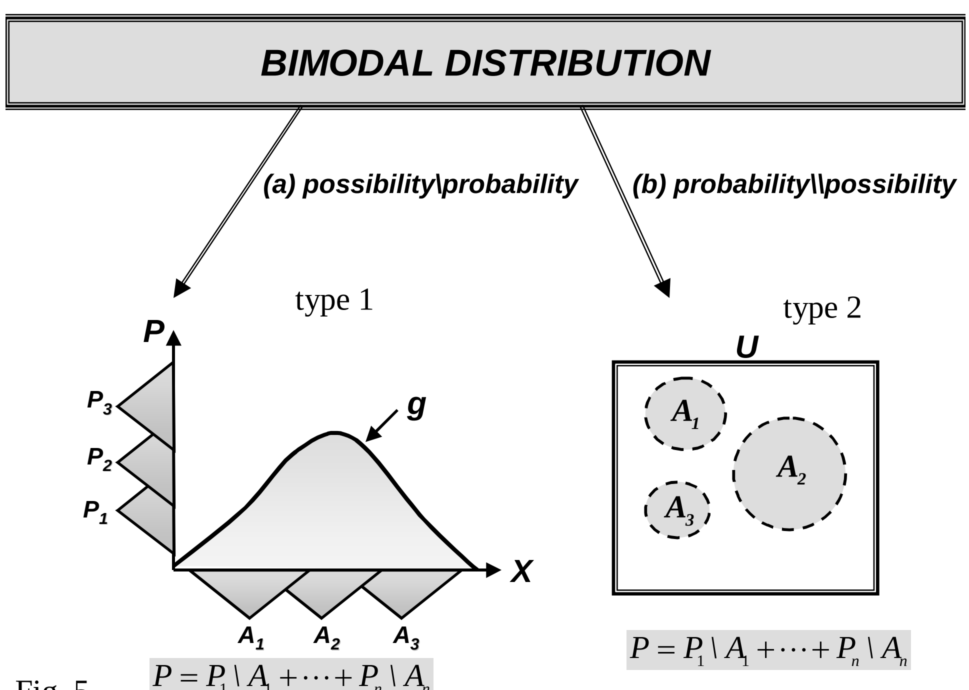


Fig. 3.



Y
$A_i \times B_i$
$R=\Sigma_i A_i \times B_i$
B_i
f
0
X
A_i
$R: \Sigma_i A_i \times B_i$

Fig. 4.



BIMODAL DISTRIBUTION
(a) possibility\probability
(b) probability\\possibility
type 1
type 2
P
P_3
P_2
P_1
g
X
A_1
A_2
A_3
U
A_1
A_2
A_3
$P = P_1 \backslash A_1 + \cdots + P_n \backslash A_n$
$P = P_1 \backslash A_1 + \cdots + P_n \backslash A_n$

Fig. 5.

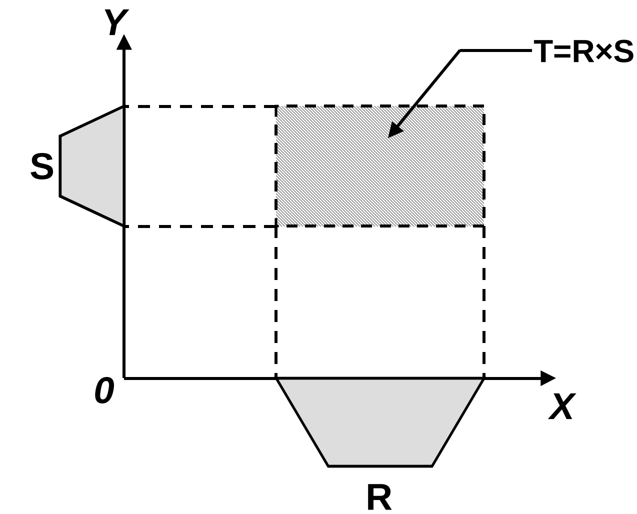


Fig. 6.

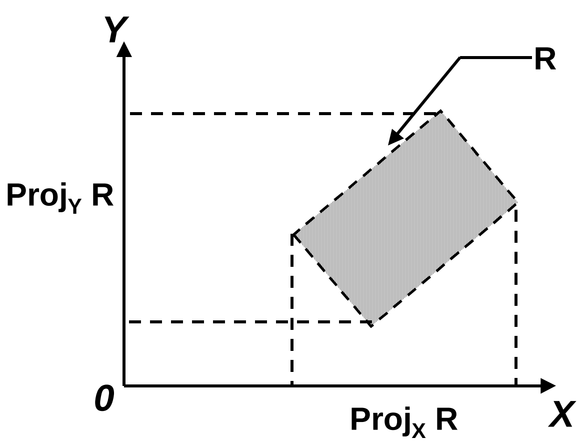


Fig. 7.

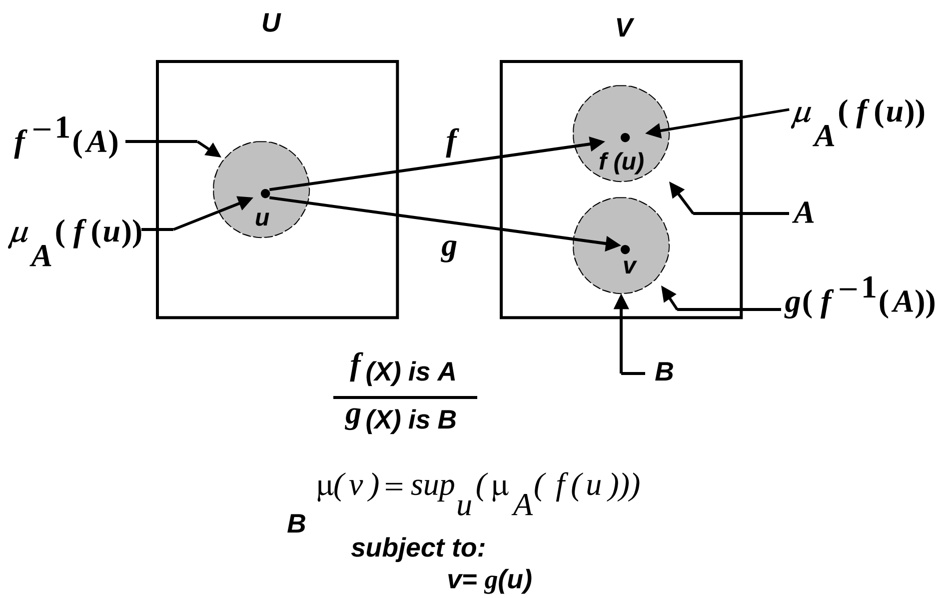
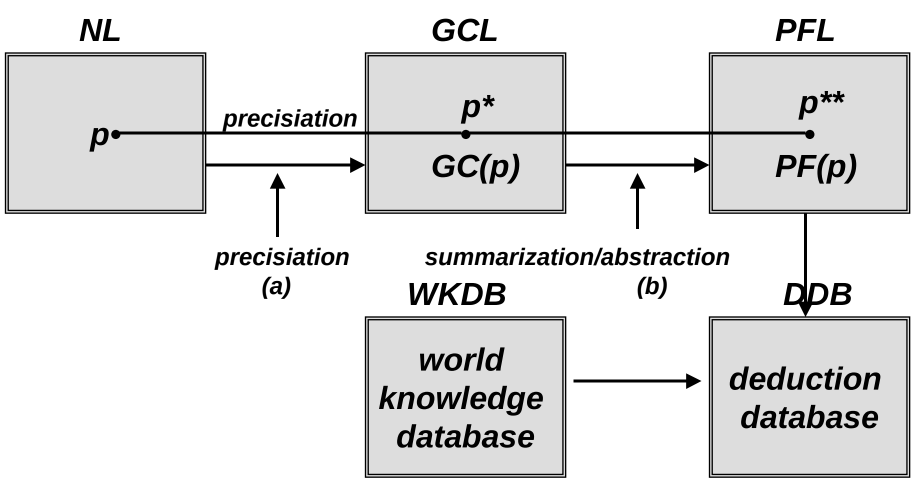
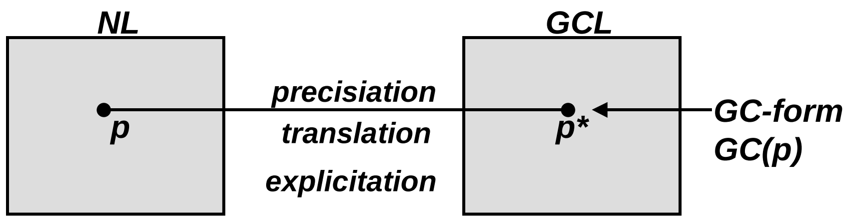


Fig. 8.



•In PNL, deduction=generalized constraint propagation
•PFL: Protoform Language
• DDB: deduction database=collection of protoformal rules governing generalized constraint propagation
• WKDB: World Knowledge Database (PNL-based)

Fig. 9.



annotated translation

p ⟶ ***X/A isr R/B*** ⟵ ***GC-form of p***

Fig. 10.

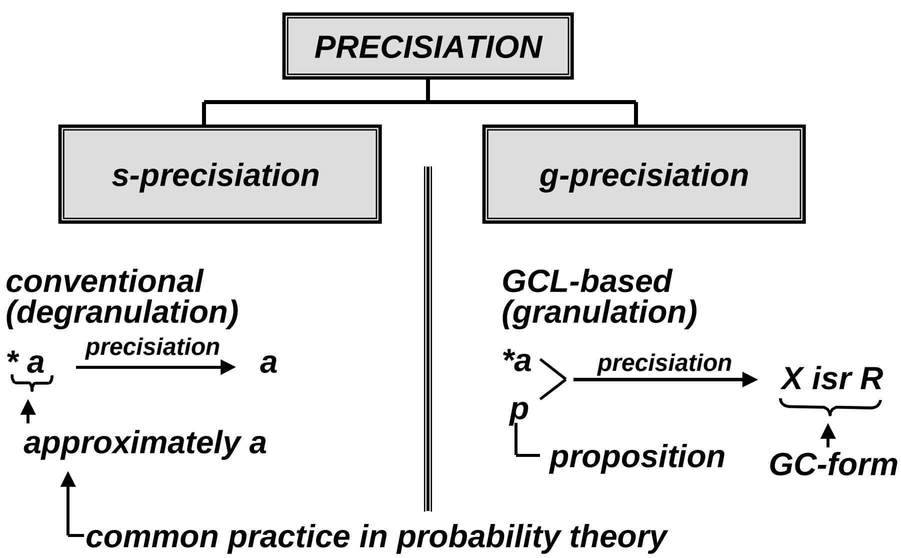


Fig. 11.

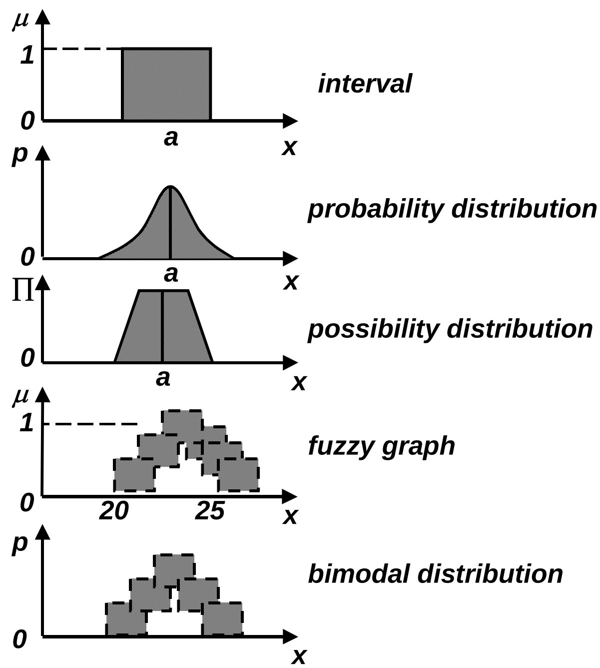

μ
1
0
a
x
interval
p
0
a
x
probability distribution
Π
0
a
x
possibility distribution
μ
1
0
20
25
x
fuzzy graph
p
0
x
bimodal distribution

Fig. 12.

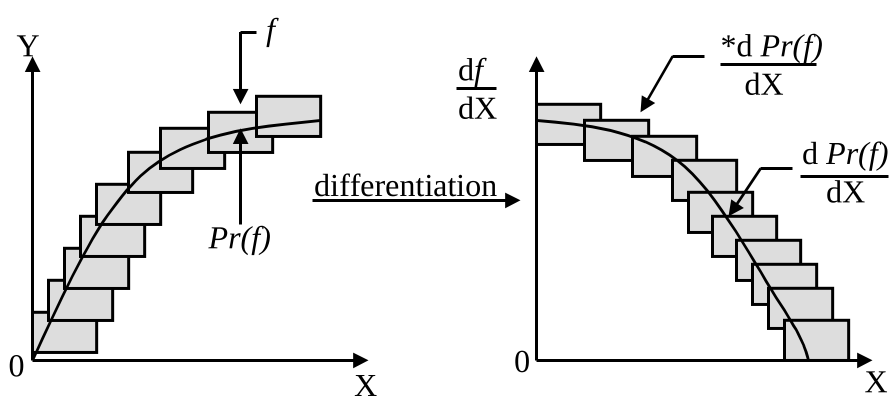

Y
f
0
X
Pr(f)
differentiation
df
dX
*d Pr(f)
dX
d Pr(f)
dX
0
X

Fig. 13.

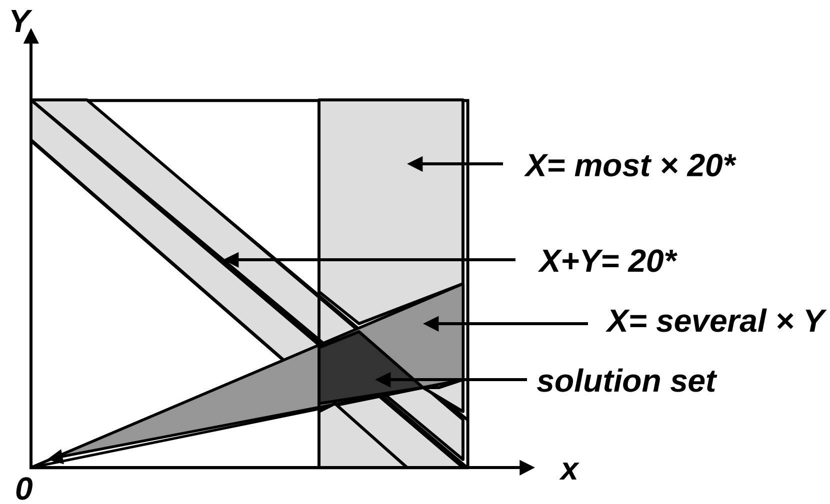

Y
X= most × 20*
X+Y= 20*
X= several × Y
solution set
x
0

Fig. 14.

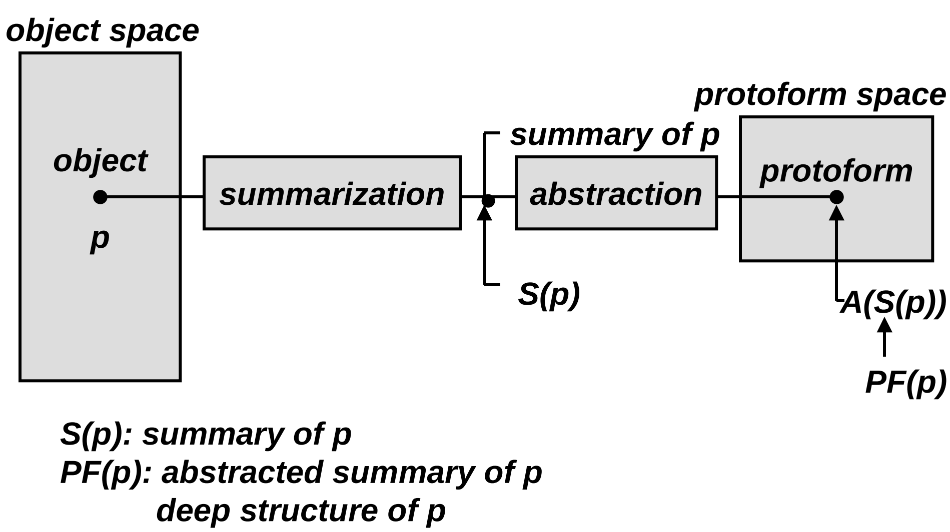
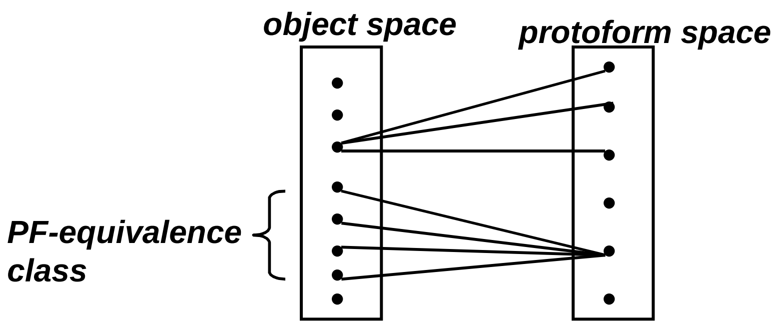

object space
object
p
summarization
summary of p
abstraction
protoform space
protoform
S(p)
A(S(p))
PF(p)
S(p): summary of p
PF(p): abstracted summary of p
deep structure of p

Fig. 15.



- ***at a given level of abstraction and summarization, objects p and q are PF-equivalent if PF(p)=PF(q)***

Fig. 16.

Alan has severe back pain. He goes to see a doctor. The doctor tells him that there are two options: (1) do nothing; and (2) do surgery. In the case of surgery, there are two possibilities: (a) surgery is successful, in which case Alan will be pain free; and (b) surgery is not successful, in which case Alan will be paralyzed from the neck down.

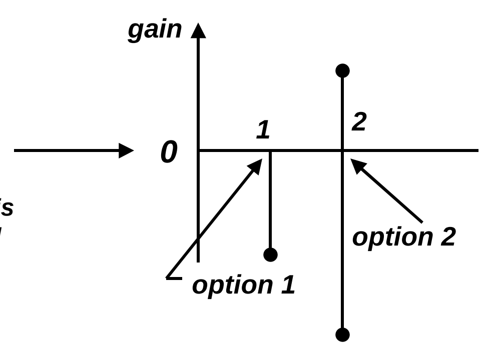


Fig. 17

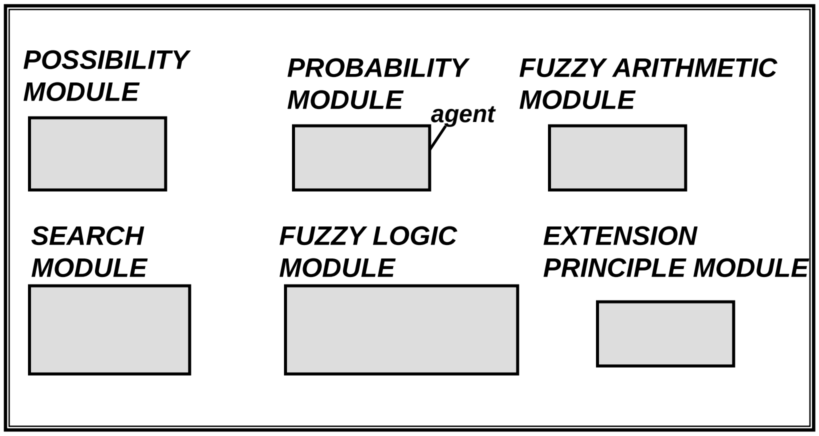

POSSIBILITY
MODULE
PROBABILITY
MODULE
agent
FUZZY ARITHMETIC
MODULE
SEARCH
MODULE
FUZZY LOGIC
MODULE
EXTENSION
PRINCIPLE MODULE

Fig. 18.

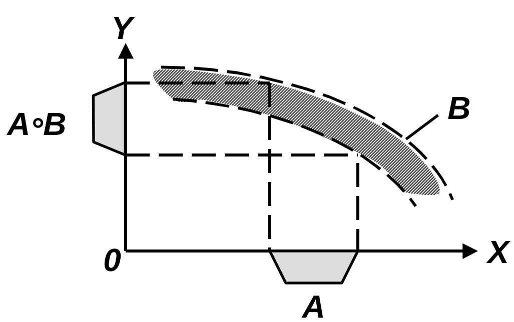

Y
A∘B
B
0
X
A

Fig. 19.

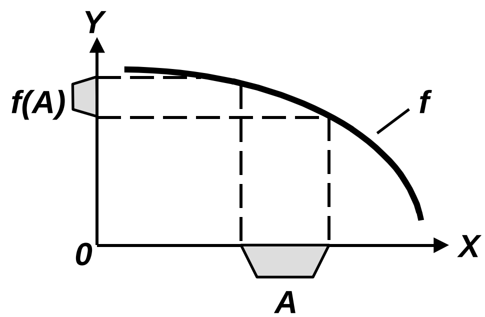
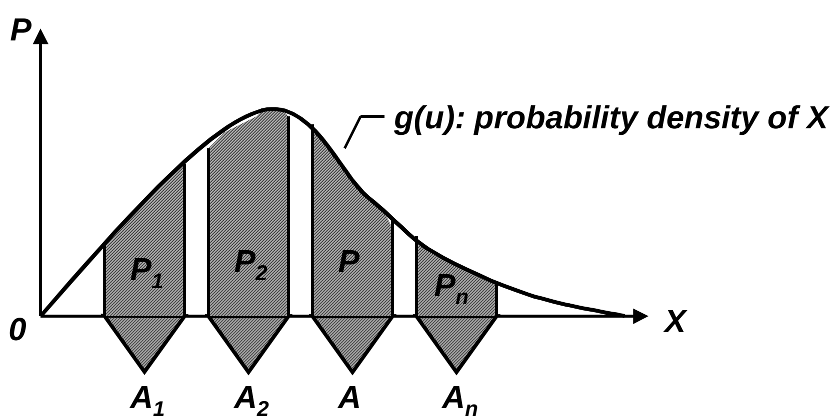


Fig. 20.



p_i is P_i : granular value of p_i , i=1, …, n
(P_i , A_i) , i=1, …, n   are given
A is given
(?P, A)

Fig. 21.

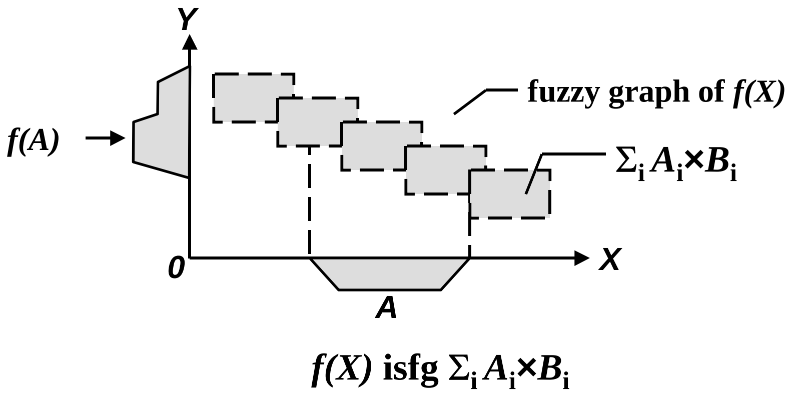

Y
fuzzy graph of f(X)
f(A)
Σi Ai×Bi
0
X
A
f(X) isfg Σi Ai×Bi

Fig. 22.

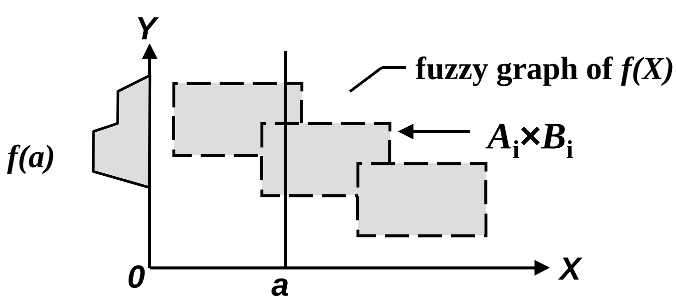

Y
fuzzy graph of f(X)
Ai×Bi
f(a)
0
a
X

Fig. 23.

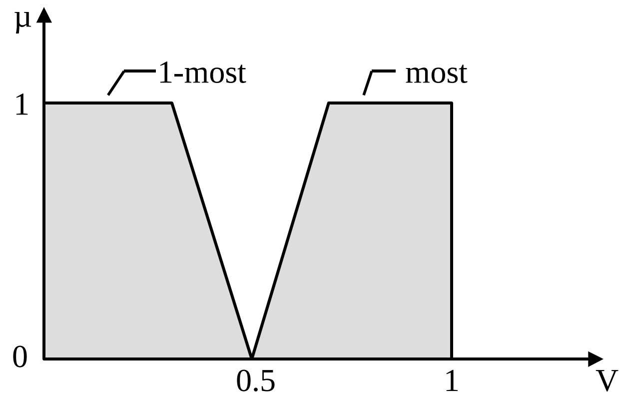


Fig. 24.