From Search Engines to Question-Answering Systems—The Problems of World Knowledge, Relevance and Deduction

Lotfi A. Zadeh

Computer Science Division
Department of EECS
UC Berkeley

June 16, 2005
WSEAS Fuzzy Systems
Lisbon, Portugal

URL: http://www-bisc.cs.berkeley.edu
URL: http://zadeh.cs.berkeley.edu/
Email: Zadeh@eecs.berkeley.edu

BACKDROP

KEY ISSUE—DEDUCTION CAPABILITY

Existing search engines, with Google at the top, have many truly remarkable capabilities. Furthermore, constant progress is being made in improving their performance. But what should be realized is that existing search engines do not have an important capability—deduction capability—the capability to synthesize an answer to a query by drawing on bodies of information which reside in various parts of the knowledge base.

CONTINUED

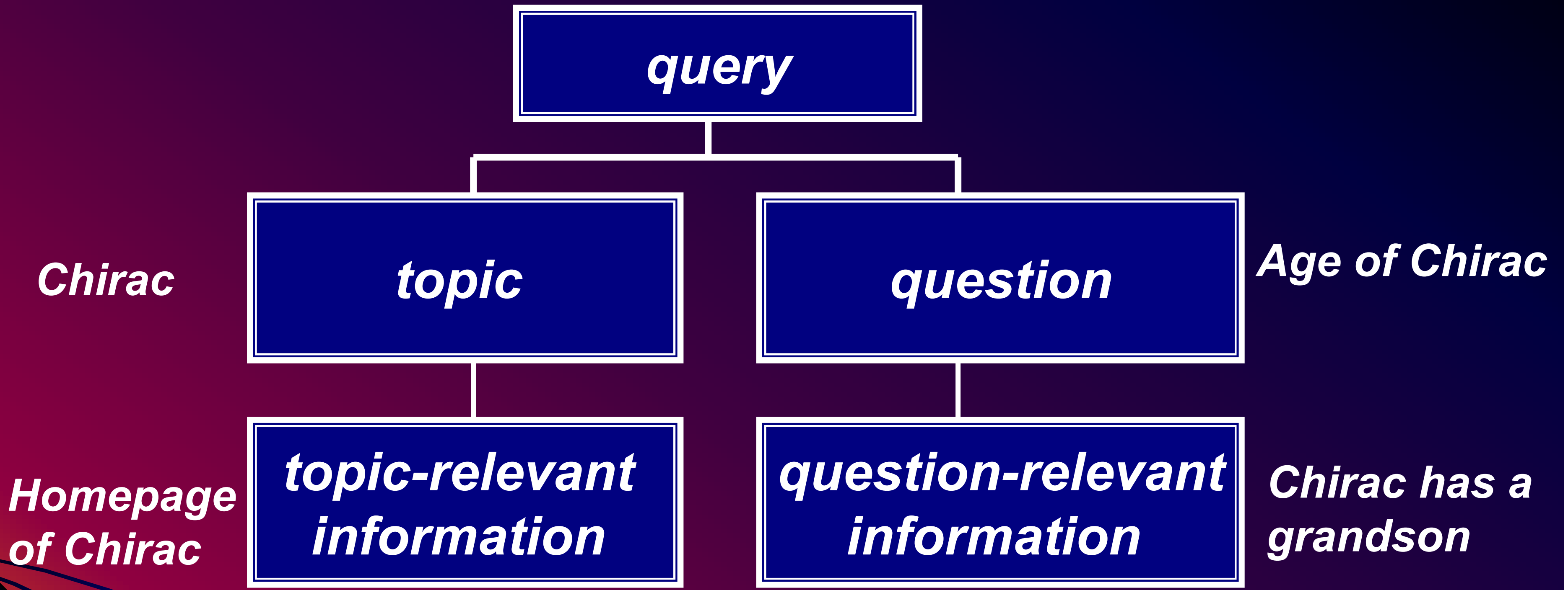
- *What should be noted, however, is that there are many widely used special purpose question-answering systems which have limited deduction capability. Examples of such systems are driving direction systems, reservation systems, diagnostic systems and specialized expert systems, especially in the domain of medicine.*

SEARCH VS. QUESTION-ANSWERING

- *A question-answering system may be viewed as a system which mechanizes question-answering*

- *A search engine in a system which partially mechanizes question-answering*

PARTIAL MECHANIZATION



- *A search engine is primarily a provider of topic relevant information*
- *User of a search engine exploits this capability to derive an answer to a question*

COMPLEXITY OF UPGRADING

- *Addition of deduction capability to a search engine is a highly complex problem—a problem which is a major challenge to computer scientists and logicians*
- *A view which is articulated in the following is that the challenge cannot be met through the use of existing methods—methods which are based on bivalent logic and probability theory*
- *To add deduction capability to a search engine it is necessary to (a) generalize bivalent logic; (b) generalize probability theory*

HISTORICAL NOTE

- 1970-1980 was a period of intense interest in question-answering and expert systems
- There was no discussion of search engines

Example: L.S. Coles, “Techniques for Information Retrieval Using an Inferential Question-Answering System with Natural Language Input,” SRI Report, 1972

- M. Nagao, J. Tsujii: Mechanism of Deduction in a Question-Answering System with Natural Language Inputd. IJCAI 1973: 285-290.
- J. R. McSkimin, J. Minker: The Use of a Semantic Network in a Deductive Question- Answering System. IJCAI 1977: 50-58.
- A. R. Aronson, B. E. Jacobs, J. Minker: A Note on Fuzzy Deduction. J. ACM 27(4): 599-603 (1980)
- W.J.H.J. Bronnenberg, H.C. Bunt, S.P.J. Lendsbergen, R.J.H. Scha, W.J. Schoenmakers and E.P.C. van Utteren. The Question Answering System PHLIQA1. In L. Bolc (editor), Natural Language Question Answering Systems. Macmillan, 1980.

GOOGLE VS. MSN ENCARTA

t_1: precisiation

q_2: What is precisiation?

r_1(Google):

[UAI] The concept of cointensive precisiation

... from data expressed in a natural language is precisiation of meaning. ...
In this perspective, the problem of precisiation is that of picking a ...

AI Magazine: Precisiated natural language

... The Concepts of Precisiability and Precisiation Language ... p is precisiable
if it can be translated into what may be called a precisiation language, ...

r_1(MSN Encarta):

Result: We couldn't find any sites containing precisiation.

SIMPLE EXAMPLES OF DEDUCTION INCAPABILITY

q_2: What is precisiation?

r_2(Google): same as r_1

r_2(MSN Encarta):

Result: We couldn't find any sites containing what is precisiation.

CONTINUED

q_1: What is the capital of New York?

q_2: What is the population of the capital of New York?

r_1(Google):

Web definitions for capital of new york Albany: state capital of New York; located in eastern New York State on the west bank of the Hudson river

News results for what is the capital of New York - View today's top stories After the twin-tower nightmare, New York is back on form, says ... - Economist - 3 hours ago
The New Raiders - BusinessWeek - 14 hours ago
Brascan acquires New York-based Hyperion Capital for $50M US

CONTINUED

r_1(MSN Encarta):

Answer:

New York, United States: Capital: Albany

CONTINUED

q_2: What is the population of the capital of New York?

r_2(Google):

News results for population of New York - View today's top stories After the twin- tower nightmare, New York is back on form, says ...
UN: World's population is aging rapidly-
New, deadly threat from AIDS virus

r_2(MSN Encarta):

MSN Encarta

Albany is the capital of New York. New York, commonly known as New York City is the largest city in New York. California surpassed New York in population in 1963.

CONTINUED

q_3: What is the distance between the largest city in Spain and the largest city in Portugal?

r_3(Google):

Porto- Oporto- Portugal Travel Planner

Munich Germany Travel Planner- Hotels Restaurants Languange ...

r_3(MSN Encarta):

ninemsn Encarta- Search View - Communism

MSN Encarta- Search View- United States (History)

MSN Encarta- Jews

CONTINUED

q_4: Age of Chirac

r_4(Google):

Jacques Chirac
Date of Birth: 29 November 1932

r_4(MSN Encarta):

... contraception and abortion, lower the voting age, and redistribute taxes. He was successful in ... and the new Gaullist prime minister, Jacques Chirac , focused on domestic matters. This arrangement ...

CONTINUED

q_5: Age of son of Chirac

r_5(Google):

... Albert, their only <u>son</u>, becomes Monaco's de facto ruler until a formal investiture
... French President Jacques <u>Chirac</u> hailed the prince's "courage and ...

r_5(MSN Encarta):

... during the Renaissance and the <u>Age</u> of Enlightenment deeply ... Corsica's most famous <u>son</u>, Napoleon Bonaparte (see Napoleon I ... In 1997 President Jacques <u>Chirac</u> lost his conservative majority in ...

CONTINUED

q_6: How many Ph.D. degrees in mathematics were granted by European Universities in 1986?

r_6(Google):

A History of the University of Podlasie

Annual Report 1996

A Brief Report on Mathematics in Iran

r_6(MSN Encarta):

Myriad

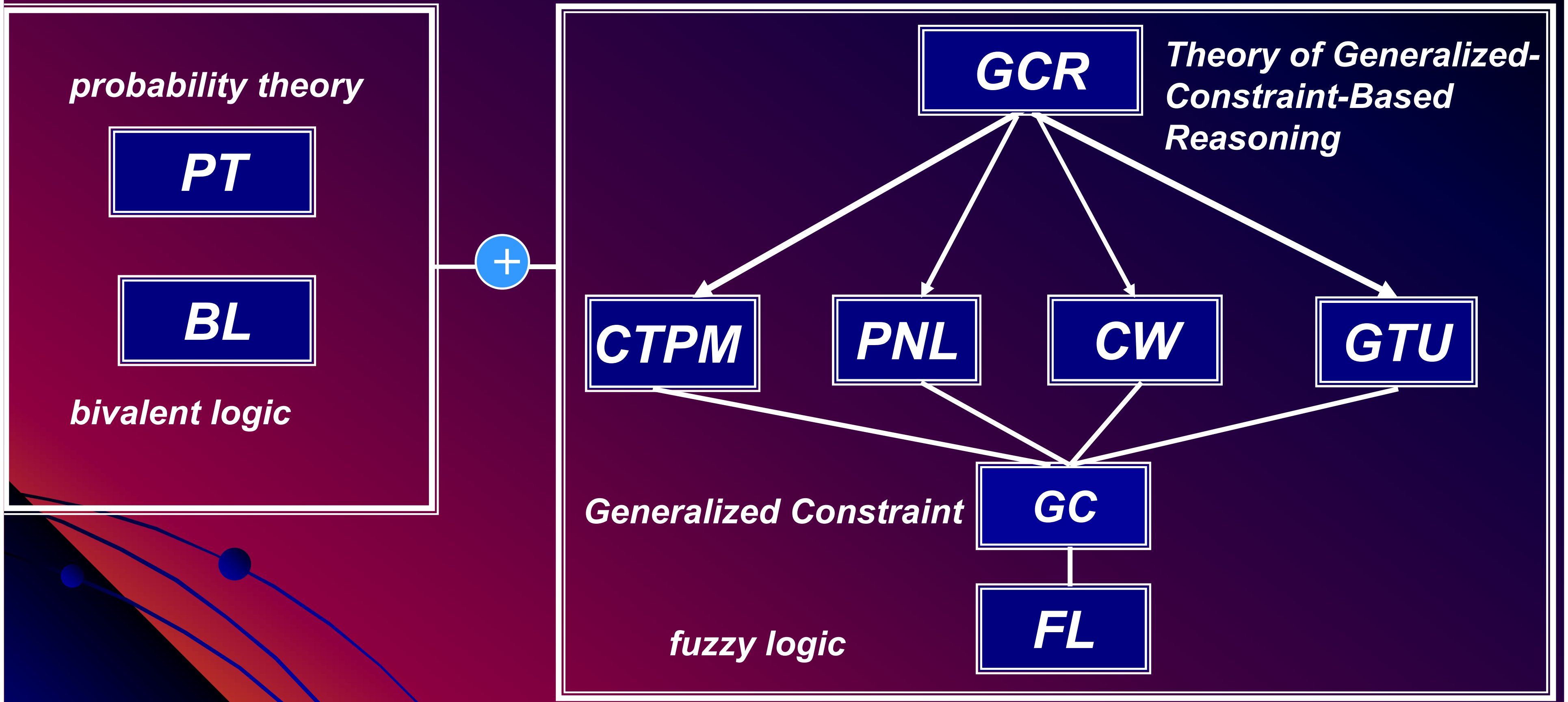
... here emerged out of many hours of discussions, over the ... 49 Master's and 3 Ph.D. degrees to Southeast Asian Americans ... the 1960s, Hmong children were granted minimal access to schooling ...

UPGRADING

- There are three major problems in upgrading a search engine to a question-answering system
    - World knowledge
    - Relevance
    - Deduction
- These problems are beyond the reach of existing methods based on bivalent logic and probability theory
- A basic underlying problem is mechanization of natural language understanding. A prerequisite to mechanization of natural language understanding is precisiation of meaning

NEED FOR NEW TOOLS

Tools in current use | New Tools



PT: standard bivalent-logic-based probability theory
CTPM : Computational Theory of Precisiation of Meaning
PNL: Precisiated Natural Language
CW: Computing with Words
GTU: Generalized Theory of Uncertainty
GCR: Theory of Generalized-Constraint-Based Reasoning

KEY CONCEPT

- ***The concept of a generalized constraint is the centerpiece of new tools—the tools that are needed to upgrade a search engine to a question-answering system***

- ***The concept of a generalized constraint serves as a bridge between linguistics and mathematics by providing a means of precisiation of propositions and concepts drawn from a natural language***

WORLD KNOWLEDGE

- World knowledge is the knowledge acquired through the experience, education and communication
  - Few professors are rich
  - There are no honest politicians
  - It is not likely to rain in San Francisco in midsummer
  - Most Swedes are tall
  - There are no mountains in Holland
  - Usually Princeton means Princeton University
  - Paris is the capital of France

COMPONENTS OF WORLD KNOWLEDGE

- *Propositional*
  - *Paris is the capital of France*
- *Conceptual*
  - *Climate*
- *Ontological*
  - *Rainfall is related to climate*
- *Existential*
  - *A person cannot have two fathers*
- *Contextual*
  - *Tall*

CONTINUED

- *Much of world knowledge is perception-based*
  - *Most Swedes are tall*
  - *Most Swedes are taller than most Italians*
  - *Usually a large house costs more than a small house*
- *Much of world knowledge is negative, i.e., relates to impossibility or nonexistence*
  - *A person cannot have two fathers*
  - *There are no honest politicians*
- *Much of world knowledge is expressed in a natural language*

PROBLEM

- *Existing methods cannot deal with deduction from perception-based knowledge*
  - *Most Swedes are tall*

    What is the average height of Swedes?

    How many are not tall?

    How many are short?
  - *A box contains about 20 black and white balls. Most are black. There are several times as many black balls as white balls.*

    How many balls are white?

THE PROBLEM OF DEDUCTION

- p_1: usually temperature is not very low
  p_2: usually temperature is not very high
  ?temperature is not very low and not very high

- most students are young
  most young students are single
  ?students are young and single

- Bryan is much older than most of his close friends
  How old is Bryan?

THE PROBLEM OF RELEVANCE

- *A major obstacle to upgrading is the concept of relevance. There is an extensive literature on relevance, and every search engine deals with relevance in its own way, some at a high level of sophistication. But what is quite obvious is that the problem of assessment of relevance is very complex and far from solution*

- *What is relevance?*
- *Relevance is not bivalent*
- *Relevance is a matter of degree, i.e., is a fuzzy concept*
- *There is no cointensive definition of relevance in the literature*

CONTINUED

Definition of relevance function

R(q/p)

- *proposition or collection of propositions* (p)
- *question or topic* (q)
- *degree of relevance of p to q* (R)

q: number of cars in California?
p: population of California is 37,000,000
To what degree is p relevant to q?

A SERIOUS COMPLICATION—NONCOMPOSITIONALITY

- $R(q/p, r) = ?$
- $R(q/p) = 0; R(q/r) = 0; R(q/p, r) \neq 0$

Example

q: How old is Mary?

p: Mary's age is the same as Carol's age

r: Carol is 32 $R(q/p) = 0; R(q/r) = 0; R(q/p, r) = 1$

- *Conclusion: relevance cannot be assessed in isolation*
- *Definition*
- *p is i-relevant to q if p is relevant to q in isolation*
- *p is i-irrelevant to q if p is not relevant to q in isolation*

RELEVANCE

- *semantic relevance*
- *statistical relevance*

q: How old is Vera

p_1: Vera has a son who is in mid-twenties

p_2: Vera has a daughter who is in mid-thirties

w: child-bearing age is about sixteen to about forty two

page ranking algorithms

word counts

keywords

MECHANIZATION OF QUESTION ANSWERING

- *Much of world knowledge and web knowledge is expressed in a natural language*
- *Natural language understanding is a prerequisite to question-answering*
- *Precisiation of meaning is a prerequisite to mechanization of natural language understanding*
- *Human natural language understanding is a prerequisite to precisiation*
- *Machines do not have the human ability to understand what has imprecise meaning*

Example: Take a few steps

THE CONCEPT OF PRECISIATION

- *The concepts of precision and imprecision have a position of centrality in science and, more generally, in human cognition. But what is not in existence is the concept of precisiation—a concept whose fundamental importance becomes apparent when we move from bivalent logic to fuzzy logic.*

WHAT IS PRECISE?

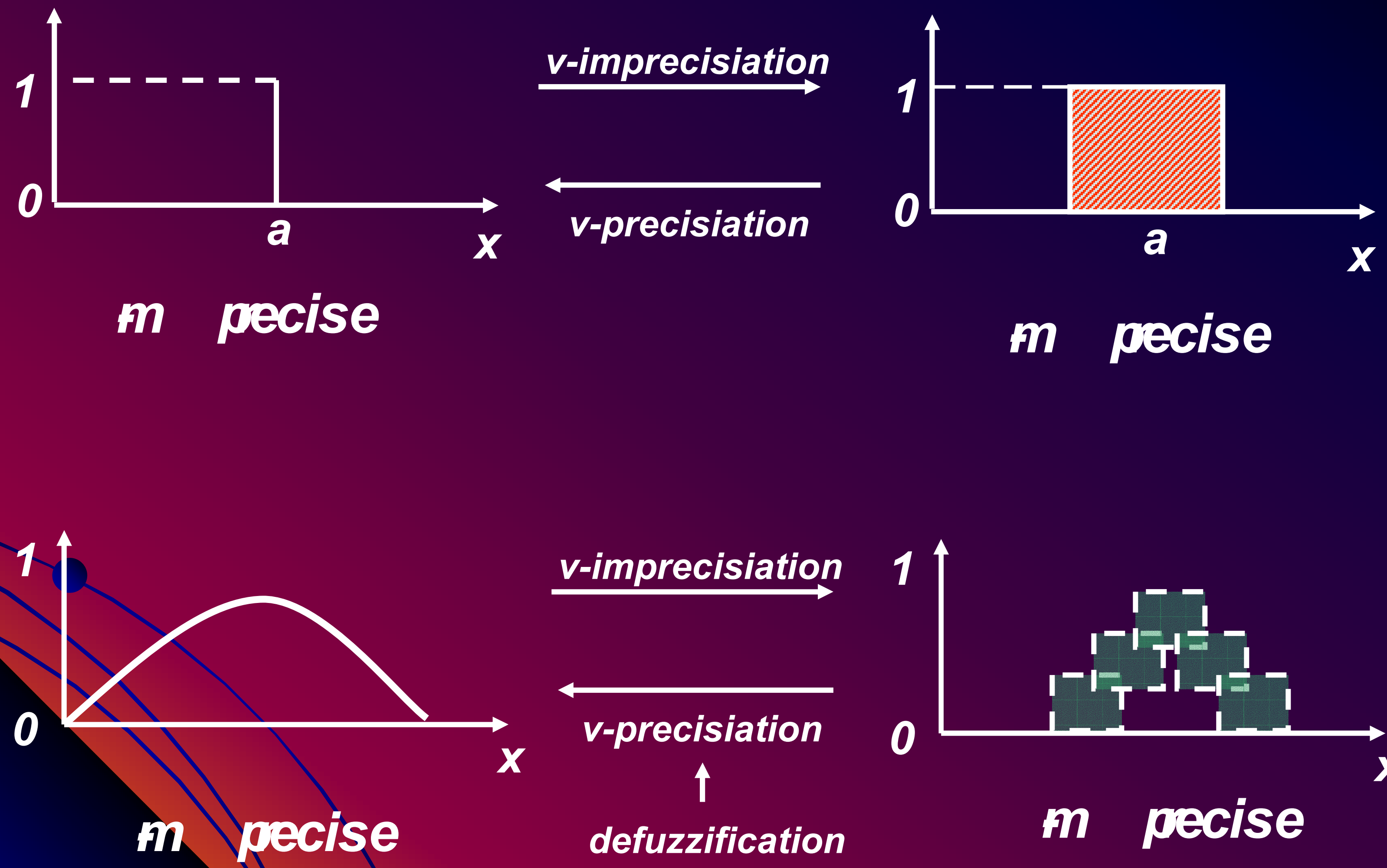
PRECISE

- **v-precise** ← precise value
- **m-precise** ← precise meaning

- p: X is a Gaussian random variable with mean m and variance σ^2. m and σ^2 are precisely defined real numbers
- p is v-imprecise and m-precise
- p: X is in the interval [a, b]. a and b are precisely defined real numbers
- p is v-imprecise and m-precise

m-precise = mathematically well-defined

PRECISIATION AND IMPRECISIATION



MODALITIES OF m-PRECISIATION

m-precisiation

- mh-precisiation ← human-oriented
- mm-precisiation ← machine-oriented

BIMODAL DICTIONARY (LEXICON) IN PNL

mh-precisiand

mm-precisiand

P

mh-precisiation

Def(p)

mm-precisiation

GC(p)

machine-oriented (mathematical)

human-oriented natural language

proposition or concept

KEY POINTS

In PNL

precisiation = mm-precisiation

- *a proposition, p, is p precisiated by representing its meaning as a generalized constraint*
- *precisiation of meaning does not imply precisiation of value*
- *"Andrea is tall" is precisiated by defining "tall" as a fuzzy set*
- *A desideratum of precisiation is cointension*
- *Informally, p and q are cointensive if the intension (attribute based meaning) of p is approximately the same as the intension (attribute-based meaning) of q*

VALIDITY OF DEFINITION

- If C is a concept and Def(C) is its definition, then Def(C) is a valid definition if it is cointensive with C

IMPORTANT CONCLUSION

- In general, cointensive, i.e., valid, definitions of fuzzy concepts cannot be formulated within the conceptual structure of bivalent logic and bivalent-logic-based probability theory
- This conclusion applies to such basic concepts as
  - Causality
  - Relevance
  - Summary
  - Intelligence
  - Mountain

PRECISIATION OF MEANING VS. UNDERSTANDING OF MEANING

- Precisiation of meaning $\neq$ Understanding of meaning
  - I understand what you said, but can you be more precise
- Use with adequate ventilation
- Unemployment is high
- Most Swedes are tall
- Most Swedes are much taller than most Italians
- Overeating causes obesity
- Causality
- Relevance
- Bear market
- Mountain
- Edge
- Approximately 5

fuzzy concepts

IMPORTANT IMPLICATION

- ***In general, a cointensive definition of a fuzzy concept cannot be formulated within the conceptual structure of bivalent logic***

To understand the meaning of this implication an analogy is helpful

ANALOGY

S		M(S)
system	modelization	model
p		**GC(p)**
proposition or concept	precisiation	precisiand

input output relation ⟶ intension

degree of match between M(S) and S ⟶ cointension

In general, it is not possible to constraint a cointensive model of a nonlinear system from linear components

PRECISIATION OF MEANING

BASIC POINT

- The meaning of a proposition, p, may be precisiated in many different ways

$$p \xrightarrow{\text{precisiation}} \left.\begin{array}{l} Pre_1(p) \\ Pre_2(p) \\ \ldots \\ Pre_n(p) \end{array}\right\} \text{precisiands of } p$$

- Conventional, bivalent-logic-based precisiation has a limited expressive power

CHOICE OF PRECISIANDS

BASIC POINT

- *The concept of a generalized constraint opens the door to an unlimited enlargement of the number of ways in which a proposition may be precisiated*
- *An optimal choice is one in which achieves a compromise between complexity and cointension*

EXAMPLE OF CONVENTIONAL DEFINITION OF FUZZY CONCEPTS

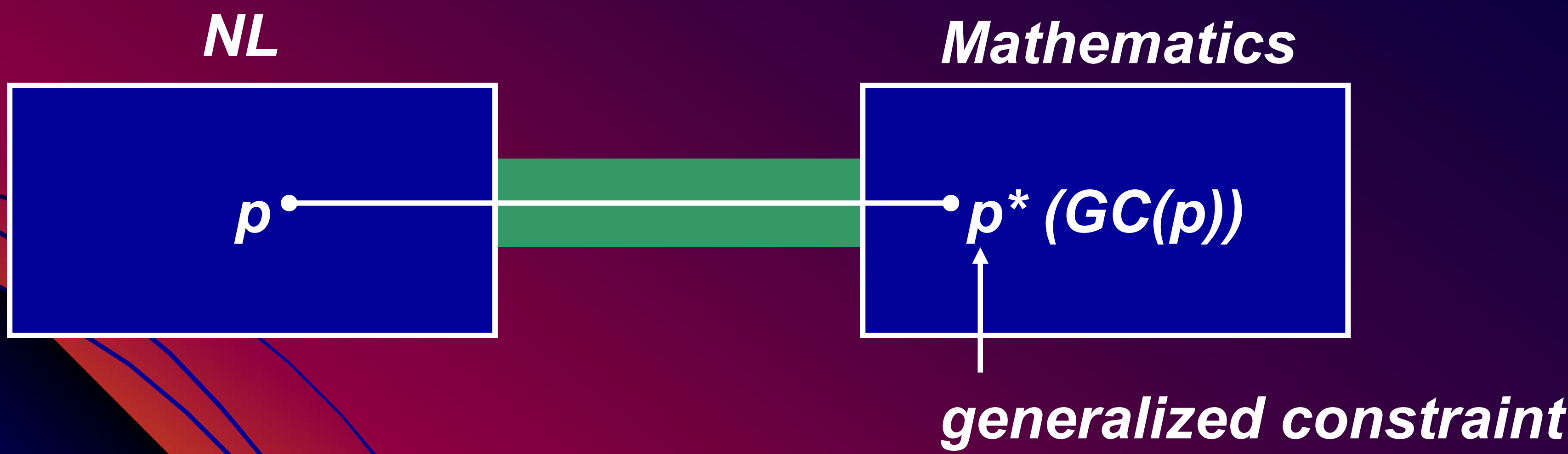
Robert Shuster (Ned Davis Research)

We classify a bear market as a 30 percent decline after 50 days, or a 13 percent decline after 145 days.

- *A problem with this definition of bear market is that it is not cointensive*

THE KEY IDEA

- *In PNL, a proposition, p, is precisiated by expressing its meaning as a generalized constraint. In this sense, the concept of a generalized constraint serves as a bridge from natural languages to mathematics.*



- *The concept of a generalized constraint is the centerpiece of PNL*

THE CONCEPT OF A GENERALIZED CONSTRAINT

GENERALIZED CONSTRAINT (Zadeh 1986)

- ***Bivalent constraint (hard, inelastic, categorical:)***

  $X \, \varepsilon \, C$ ← ***constraining bivalent relation***

- ***Generalized constraint:***

  $\boxed{X \text{ isr } R}$

  - R: ***constraining non-bivalent (fuzzy) relation***
  - r: ***index of modality (defines semantics)***
  - X: ***constrained variable***

$$r: \underbrace{\varepsilon \mid = \mid \leq \mid \geq \mid \subset \mid \ldots}_{\text{bivalent}} \mid \text{blank} \mid p \mid v \mid u \mid rs \mid fg \mid ps \mid \ldots$$

primary (blank | p | v)

CONTINUED

- *constrained variable*
  - *X is an n-ary variable, $X= (X_1, \ldots, X_n)$*
  - *X is a proposition, e.g., Leslie is tall*
  - *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 generalized constraint, X: Y isr R*
  - *X is a group variable. In this case, there is a group, $G[A]: (Name_1, \ldots, Name_n)$, with each member of the group, $Name_i$, $i =1, \ldots, n$, associated with an attribute-value, A_i. A_i may be vector-valued. Symbolically*

$$G[A]: (Name_1/A_1+\ldots+Name_n/A_n)$$

Basically, X is a relation

SIMPLE EXAMPLES

- ***"Check-out time is 1 pm," is an instance of a generalized constraint on check-out time***

- ***"Speed limit is 100km/h" is an instance of a generalized constraint on speed***

- ***"Vera is a divorcee with two young children," is an instance of a generalized constraint on Vera's age***

GENERALIZED CONSTRAINT—MODALITY r

X isr R

r: =	equality constraint: X=R is abbreviation of X is=R
r: ≤	inequality constraint: X ≤ R
r:⊂	subsethood constraint: X ⊂ R
r: blank	possibilistic constraint; X is R; R is the possibility distribution of X
r: v	veristic constraint; X isv R; R is the verity distribution of X
r: p	probabilistic constraint; X isp R; R is the probability distribution of X

Standard constraints: bivalent possibilistic, bivalent veristic and probabilistic

CONTINUED

r: rs random set constraint; X isrs R; R is the set-valued probability distribution of X

r: fg fuzzy graph constraint; X isfg R; X is a function and R is its fuzzy graph

r: u usuality constraint; X isu R means usually (X is R)

r: g group constraint; X isg R means that R constrains the attribute values of the group

PRIMARY GENERALIZED CONSTRAINTS

Possibilistic examples:

- *Monika is young* ⟶ *Age (Monika) is young*

  X R

- *Monika is much younger than Maria* ⟶

  (Age (Monika), Age (Maria)) is much younger

  X R

- *most Swedes are tall*

  ⟶ *ΣCount (tall.Swedes/Swedes) is most*

  R X

STANDARD CONSTRAINTS

- ***Bivalent possibilistic: X ε C (crisp set)***
- ***Bivalent veristic: Ver(p) is true or false***
- ***Probabilistic: X isp R***
- ***Standard constraints are instances of generalized constraints which underlie methods based on bivalent logic and probability theory***

EXAMPLES: PROBABILISITIC

- X is a normally distributed random variable with mean m and variance σ^2 $\longrightarrow$

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

- X is a random variable taking the values u_1, u_2, u_3 with probabilities p_1, p_2 and p_3, respectively $\longrightarrow$

  X isp $(p_1\backslash u_1+p_2\backslash u_2+p_3\backslash u_3)$

EXAMPLES: VERISTIC

- *Robert is half German, quarter French and quarter Italian*

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

- *Robert resided in London from 1985 to 1990*

  Reside (Robert, London) isv [1985, 1990]

GENERALIZED CONSTRAINT—SEMANTICS

A generalized constraint, GC, is associated with a test-score function, ts(u), 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, it may be an element of a semi-ring, a lattice, or more generally, a partially ordered set, or a bimodal distribution.

example: possibilistic constraint, X is R

$$X \text{ is } R \longrightarrow Poss(X=u) = \mu_R(u)$$

$$ts(u) = \mu_R(u)$$

TEST-SCORE FUNCTION

- *GC(X): generalized constraint on X*
- *X takes values in U= {u}*
- *test-score function ts(u): degree to which u satisfies GC*
- *ts(u) may be defined (a) directly (extensionally) as a function of u; or indirectly (intensionally) as a function of attributes of u*

  intensional definition=attribute based definition

- *example (a) Andrea is tall 0.9*

  (b) Andrea's height is 175cm; $\mu_{tall}(175)=0.9$; Andrea is 0.9 tall

CONSTRAINT QUALIFICATION

- *p isr R means r- value of p is R*

in particular

p isp R ⟶ Prob(p) is R (probability qualification)

p isv R ⟶ Tr(p) is R (truth (verity) qualification)

p is R ⟶ Poss(p) is R (possibility qualification)

examples

(X is small) isp likely ⟶ Prob{X is small} is likely

(X is small) isv very true ⟶ Ver{X is small} is very true

(X isu R) ⟶ Prob{X is R} is usually

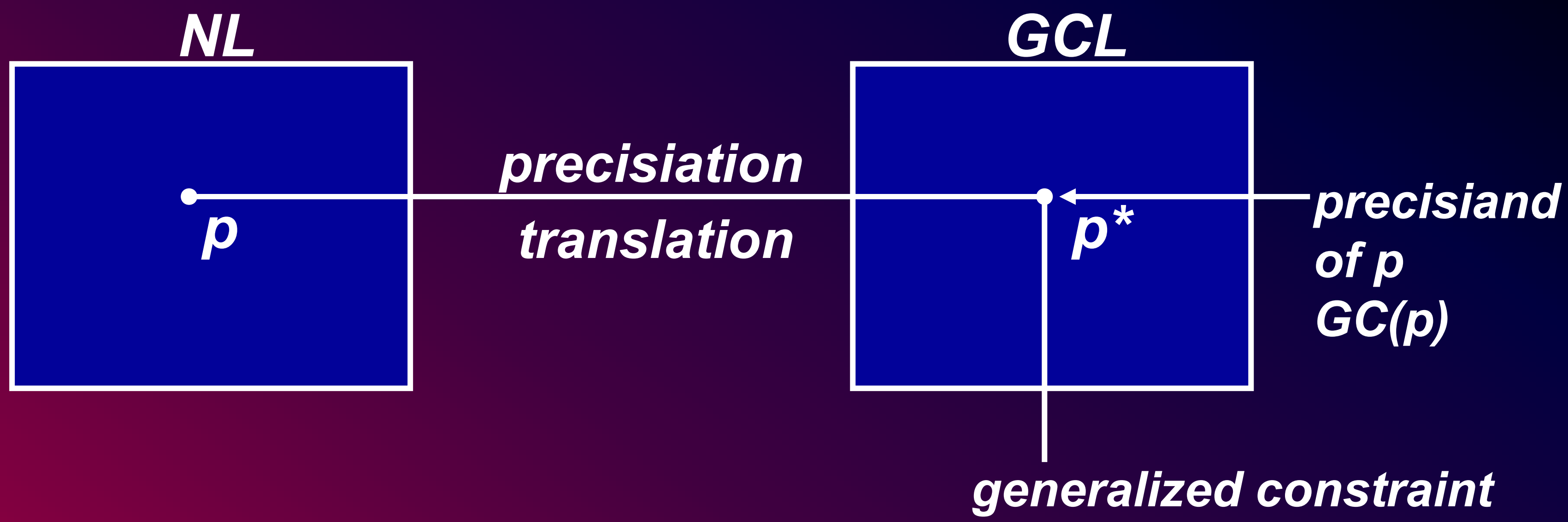
STANDARD CONSTRAINT LANGUAGE (SCL)

- SCL is a subset of GCL

- SCL is generated by combination, qualification and propagation of standard constraints

PRECISIATION = TRANSLATION INTO GCL

BASIC STRUCTURE



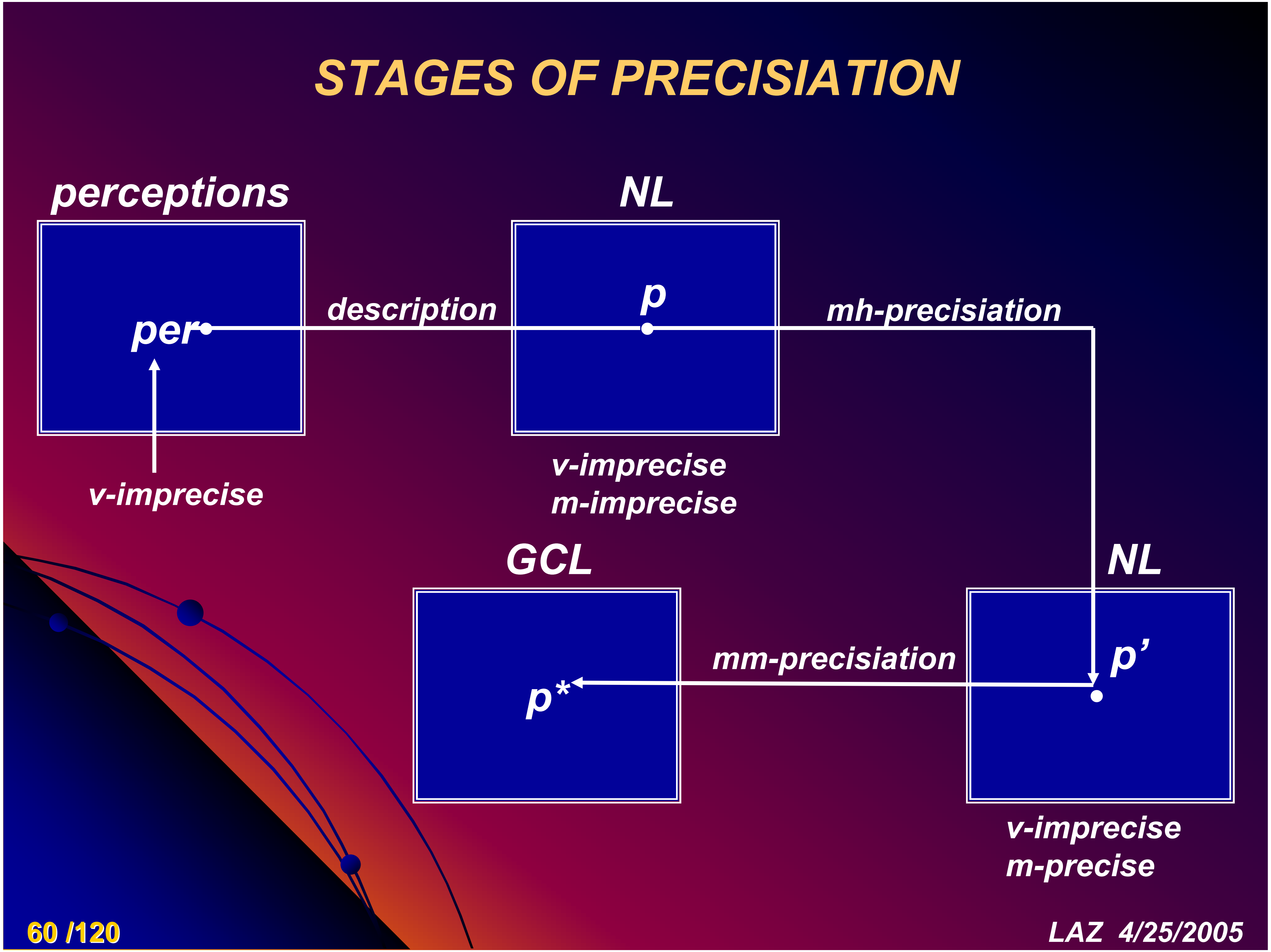
annotation

p ⟶ *X/A isr R/B* ⟵ *GC-form of p*

example

p: Carol lives in a small city near San Francisco

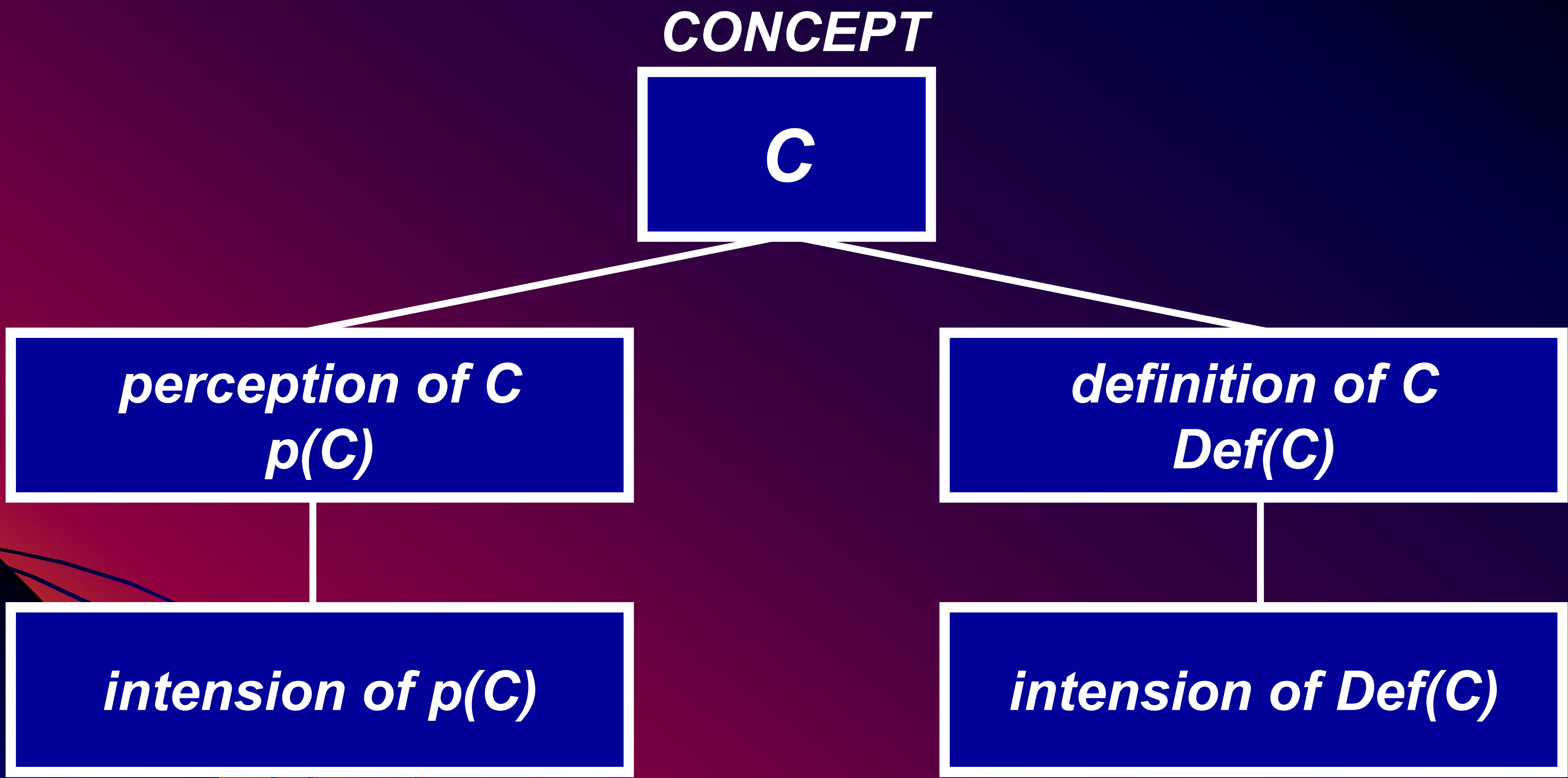
X/Location(Residence(Carol)) is R/NEAR[City] $\wedge$ *SMALL[City]*


STAGES OF PRECISIATION
perceptions
NL
per
description
p
mh-precisiation
v-imprecise
v-imprecise
m-imprecise
GCL
NL
p*
mm-precisiation
p'
v-imprecise
m-precise

COINTENSIVE PRECISIATION

- ***In general, precisiand of p is not unique. If $GC_1(p)$, …, $GC_n(p)$ are possible precisiands of p, then a basic question which arises is: which of the possible precisiands should be chosen to represent the meaning of p? There are two principal criteria which govern the choice: (a) Simplicity and (b) Cointension. Informally, the cointension of $GC_i(p)$, I=1, …, n, is the degree to which the meaning of $GC_i(p)$ approximates to the intended meaning of p. More specifically, $GC_i(p)$ is coextensive with p, or simply coextensive, if the degree to which the intension of $GC_i(p)$, with intension interpreted in its usual logical sense, approximates to the intended intension of p.***

COINTENSION OF DEFINITION

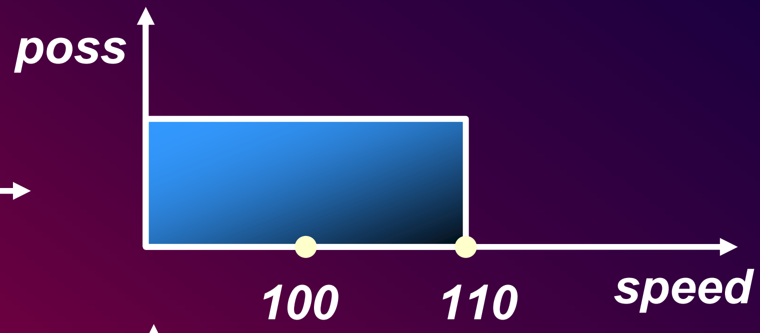


cointension: degree of goodness of fit of the intension of definiens to the intension of definiendum

EXAMPLE

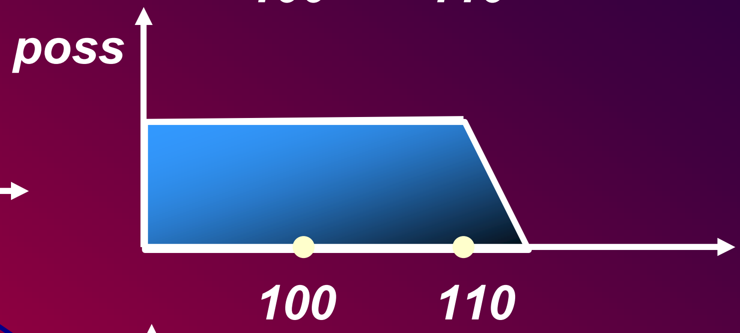
- p: Speed limit is 100 km/h

p →



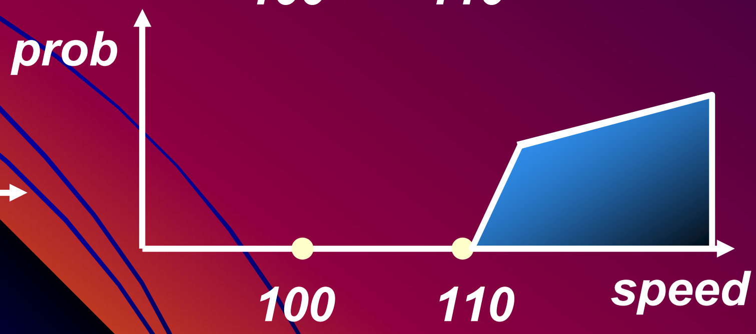
cg precisiation
r = blank (possibilistic)

p →



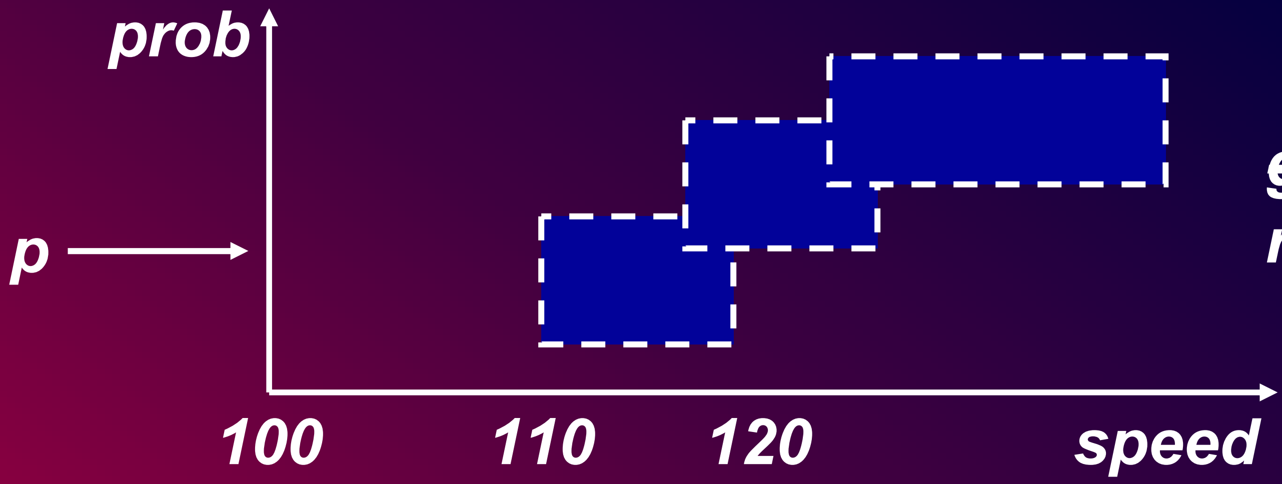
g precisiation
r = blank (possibilistic)

p →



g precisiation
r = p (probabilistic)

CONTINUED



g precisiation
r = bm (bimodal)

*If Speed is less than *110, Prob(Ticket) is low*

*If Speed is between *110 and *120, Prob(Ticket) is medium*

*If Speed is greater than *120, Prob(Ticket) is high*

PRECISIATION

s-precisiation

conventional (degranulation)

$\underbrace{*a}_{\text{approximately } a} \xrightarrow{\text{precisiation}} a$

approximately a

common practice in probability theory

g-precisiation

GCL-based (granulation)

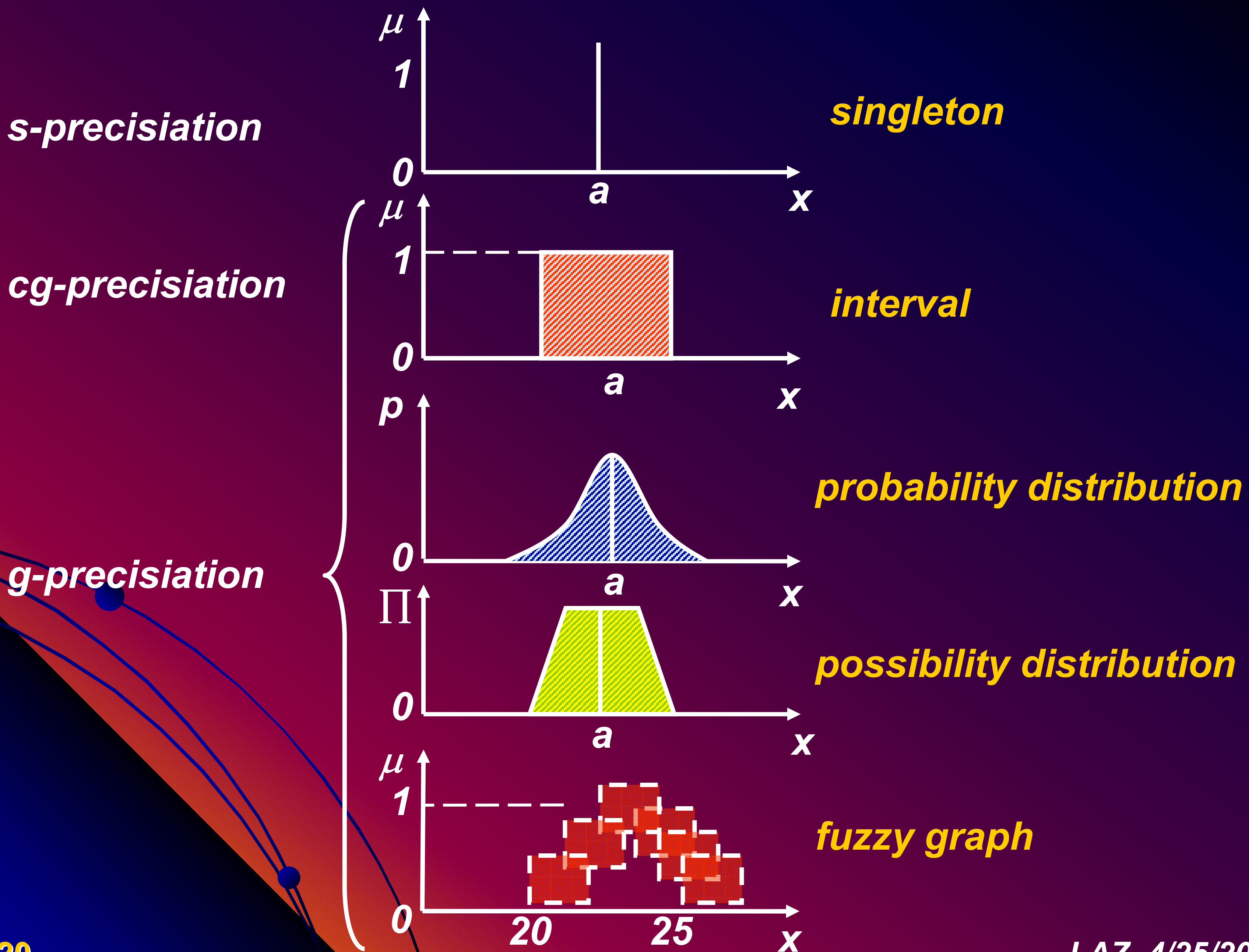
*a, p $\xrightarrow{\text{precisiation}}$ $\underbrace{X \text{ isr } R}_{\text{GC-form}}$

p: proposition

GC-form

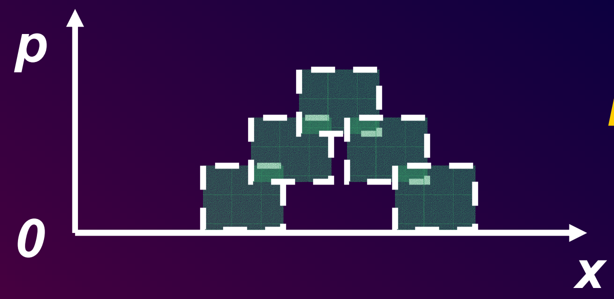
- cg-precisiation: crisp granular precisiation

PRECISIATION OF "approximately a," *a



CONTINUED

g-precisiation



bimodal distribution

GCL-based (maximal generality)

$$*a \xrightarrow{\text{g-precisiation}} \underbrace{X \ isr \ R}_{\text{GC-form}}$$

KEY POINT

- *A major limitation of bivalent-logic-based methods of concept definition is their intrinsic inability to lead to cointensive definitions of fuzzy concepts, that is concepts which are a matter of degree. Such concepts are pervassive in human knowledge and cognition.*

Examples:

- *Causality*
- *Relevance*
- *Summary*
- *Mountain*
- *Edge*
- *Pornography*

RELEVANCE AND DEDUCTION

VERA'S AGE

- *q: How old is Vera?*
- *p_1: Vera has a son, in mid-twenties*
- *p_2: Vera has a daughter, in mid-thirties*
- *wk: the child-bearing age ranges from about 16 to about 42*

CONTINUED

$R(q/p_1, p_2, wk)$: $\Pi_a = \geq \circ$ *51 $\cap$ $\leq \circ$ *67

*a: approximately a How is *a defined?

PRECISIATION AND DEDUCTION

- *p: most Swedes are tall*

  p: ΣCount(tall.Swedes/Swedes) is most*

further precisiation

h(u): height density function

h(u)du: fraction of Swedes whose height is in [u, u+du], $a \leq u \leq b$

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

CONTINUED

- $\Sigma Count(tall.Swedes/Swedes) = \int_a^b h(u)\mu_{tall}(u)du$

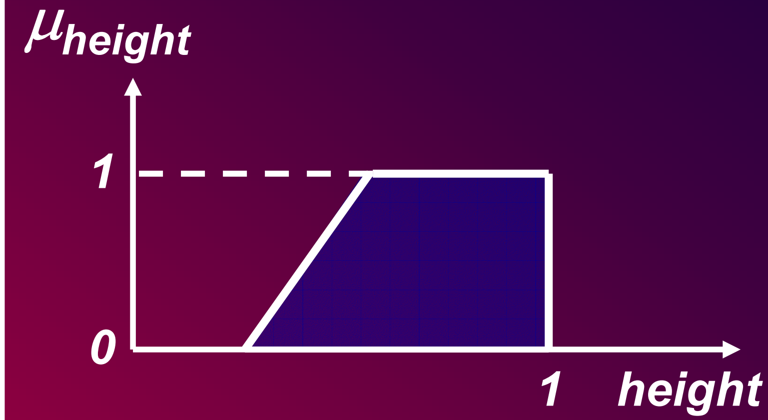
- constraint on h

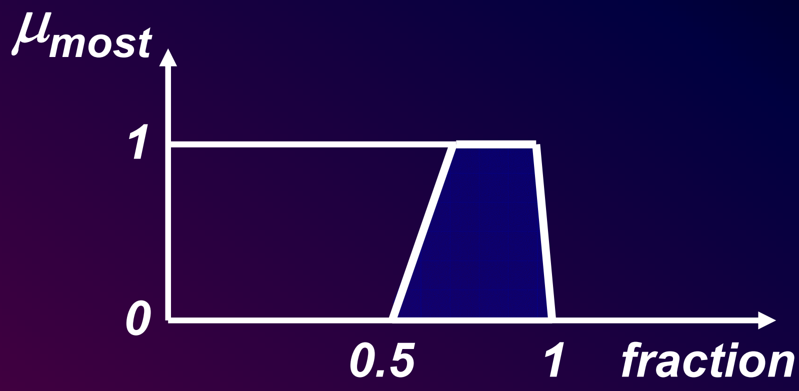
$\int_a^b h(u)\mu_{tall}(u)du$ is most

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

CALIBRATION / PRECISIATION

- *calibration*





- *precisiation*

most Swedes are tall ⟶ $\pi(h) = \mu_{most}\left(\int_a^b h(u)\mu_{tall}(u)du\right)$

h: count density function

- *Frege principle of compositionality—precisiated version*
- *precisiation of a proposition requires precisiations (calibrations) of its constituents*

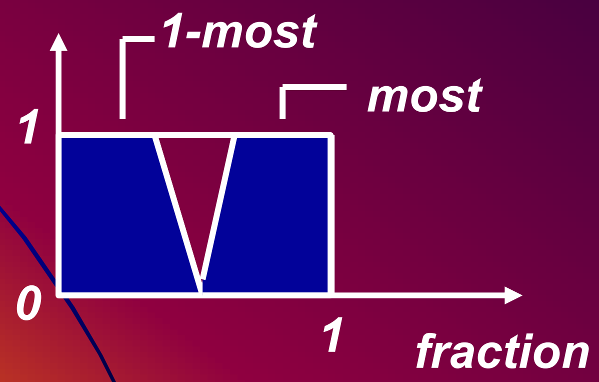
DEDUCTION

q: How many Swedes are not tall

q*: $\boxed{\int_a^b h(u)\mu_{not.tall}(u)du}$ is ? Q

solution: $\boxed{\int_a^b h(u)(1-\mu_{tall}(u))du =}$

$$\boxed{\int_a^b h(u)du - \int_a^b h(u)\mu_{tall}(u)du = 1 - most}$$



DEDUCTION

q: How many Swedes are short

q:* $\int_a^b h(u)\mu_{short}(u)du$ *is ? Q*

solution: $\int_a^b h(u)\mu_{tall}(u)$ *is most*

$\int_a^b h(u)\mu_{short}(u)$ *is ? Q*

extension principle

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

subject to

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

CONTINUED

q: What is the average height of Swedes?

q:* $\int_a^b h(u)udu$ *is ? Q*

solution: $\dfrac{\int_a^b h(u)\mu_{tall}(u)du \text{ is most}}{\int_a^b h(u)udu \quad \text{is ? Q}}$

extension principle

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

subject to

$$v = \int_a^b h(u)udu$$

PROTOFORM LANGUAGE

PFL

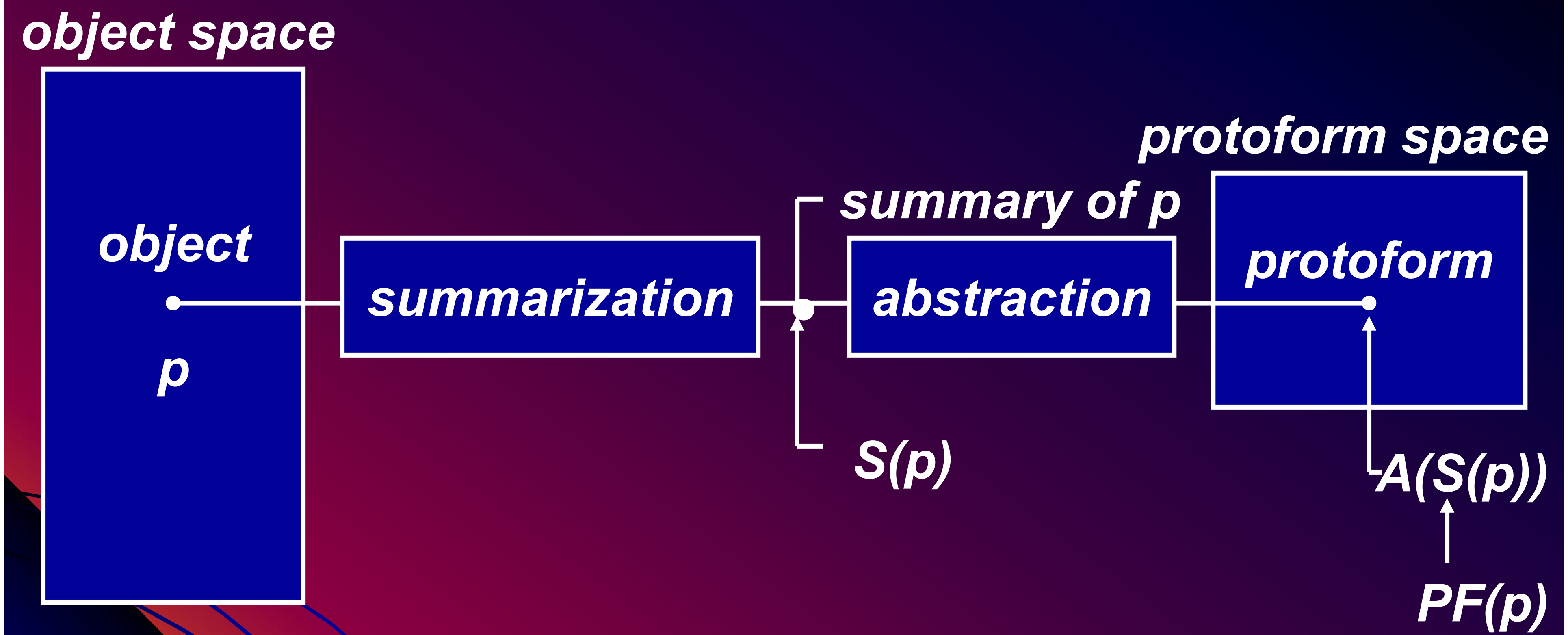
THE CONCEPT OF A PROTOFORM

PREAMBLE

- *As we move further into the age of machine intelligence and automated reasoning, a daunting problem becomes harder and harder to master. How can we cope with the explosive growth in knowledge, information and data. How can we locate and infer from decision-relevant information which is embedded in a large database.*

  Among the many concepts that relate to this issue there are four that stand out in importance: organization, representation, search and deduction. In relation to these concepts, a basic underlying concept is that of a protoform—a concept which is centered on the confluence of abstraction and summarization

CONTINUED



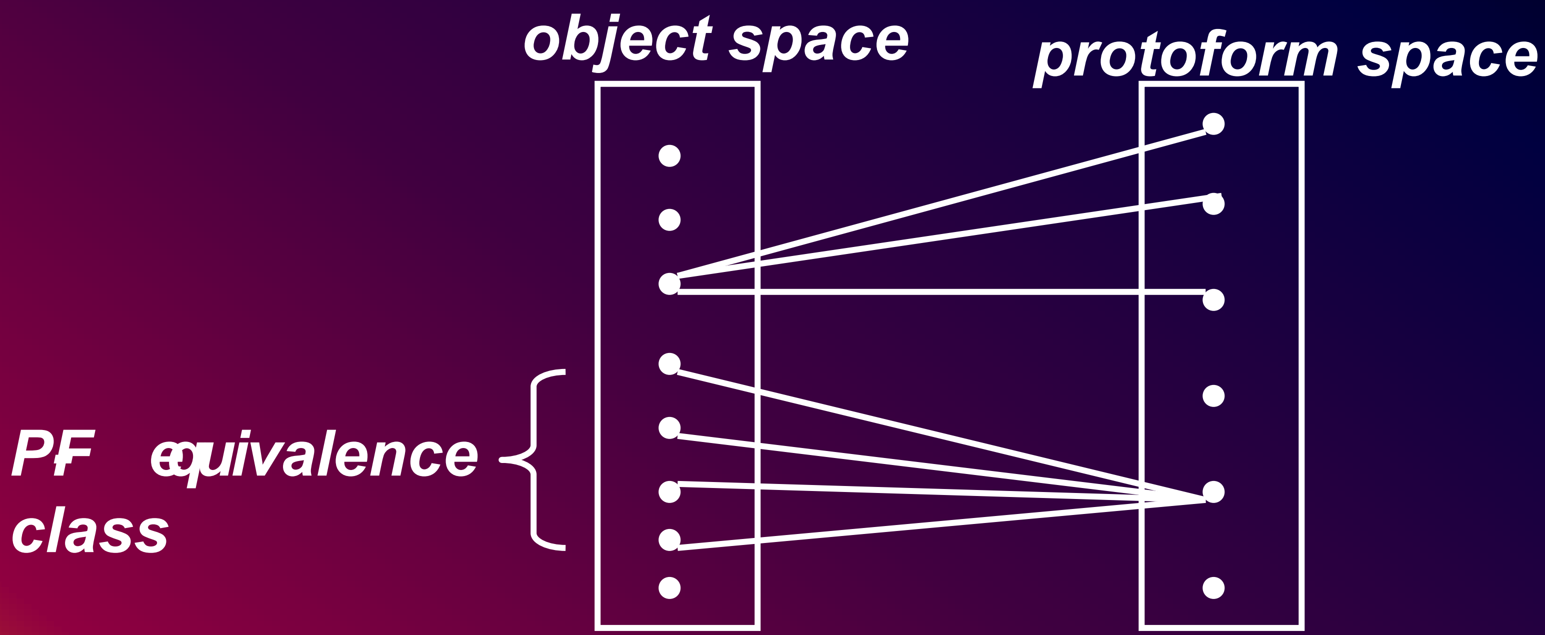
PF(p): abstracted summary of p
deep structure of p

- *protoform equivalence*
- *protoform similarity*

WHAT IS A PROTOFORM?

- *protoform = abbreviation of prototypical form*
- *informally, a protoform, A, of an object, B, written as A=PF(B), is an abstracted summary of B*
- *usually, B is lexical entity such as proposition, question, command, scenario, decision problem, etc*
- *more generally, B may be a relation, system, geometrical form or an object of arbitrary complexity*
- *usually, A is a symbolic expression, but, like B, it may be a complex object*
- *the primary function of PF(B) is to place in evidence the deep semantic structure of B*

PROTOFORMS



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

example

p: Most Swedes are tall	*Count (A/B) is Q*
q: Few professors are rich	*Count (A/B) is Q*

EXAMPLES

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

instantiation

abstraction

- Monika is much younger than Robert ⟶
  (Age(Monika), Age(Robert) is much.younger ⟶
  D(A(B), A(C)) is E
- Usually Robert returns from work at about 6:15pm ⟶
  Prob{Time(Return(Robert)} is 6:15*} is usually ⟶
  Prob{A(B) is C} is D

usually

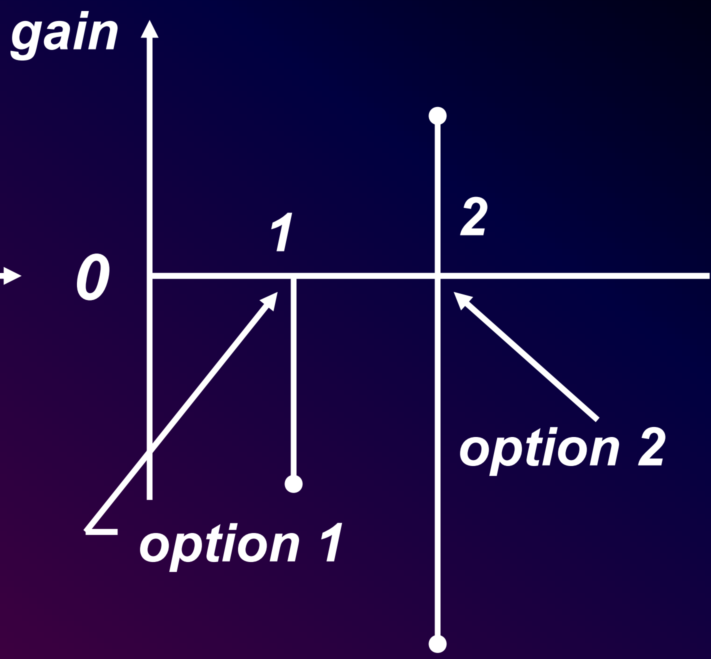
6:15*

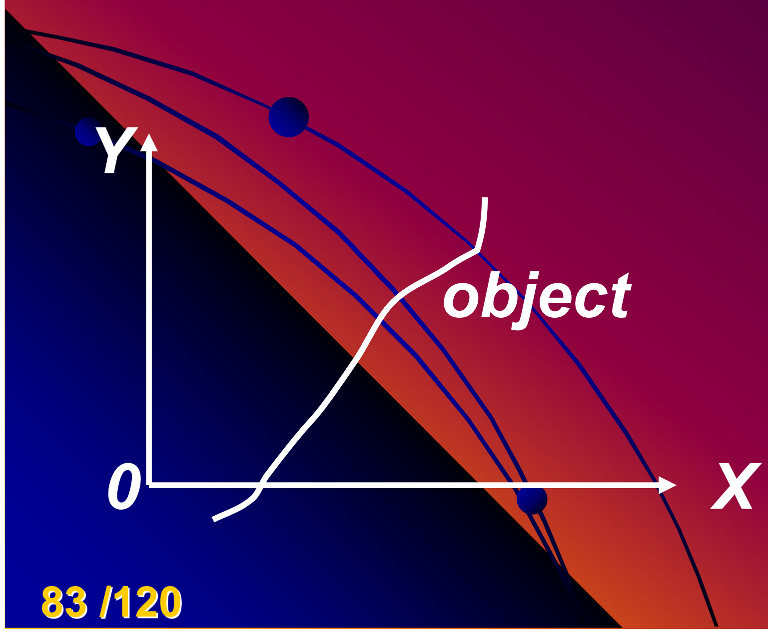
Return(Robert)

Time

EXAMPLES

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. Question: Should Alan elect surgery?





PROTOFORMAL SEARCH RULES

example

query: What is the distance between the largest city in Spain and the largest city in Portugal?

protoform of query: ?Attr (Desc(A), Desc(B))

procedure

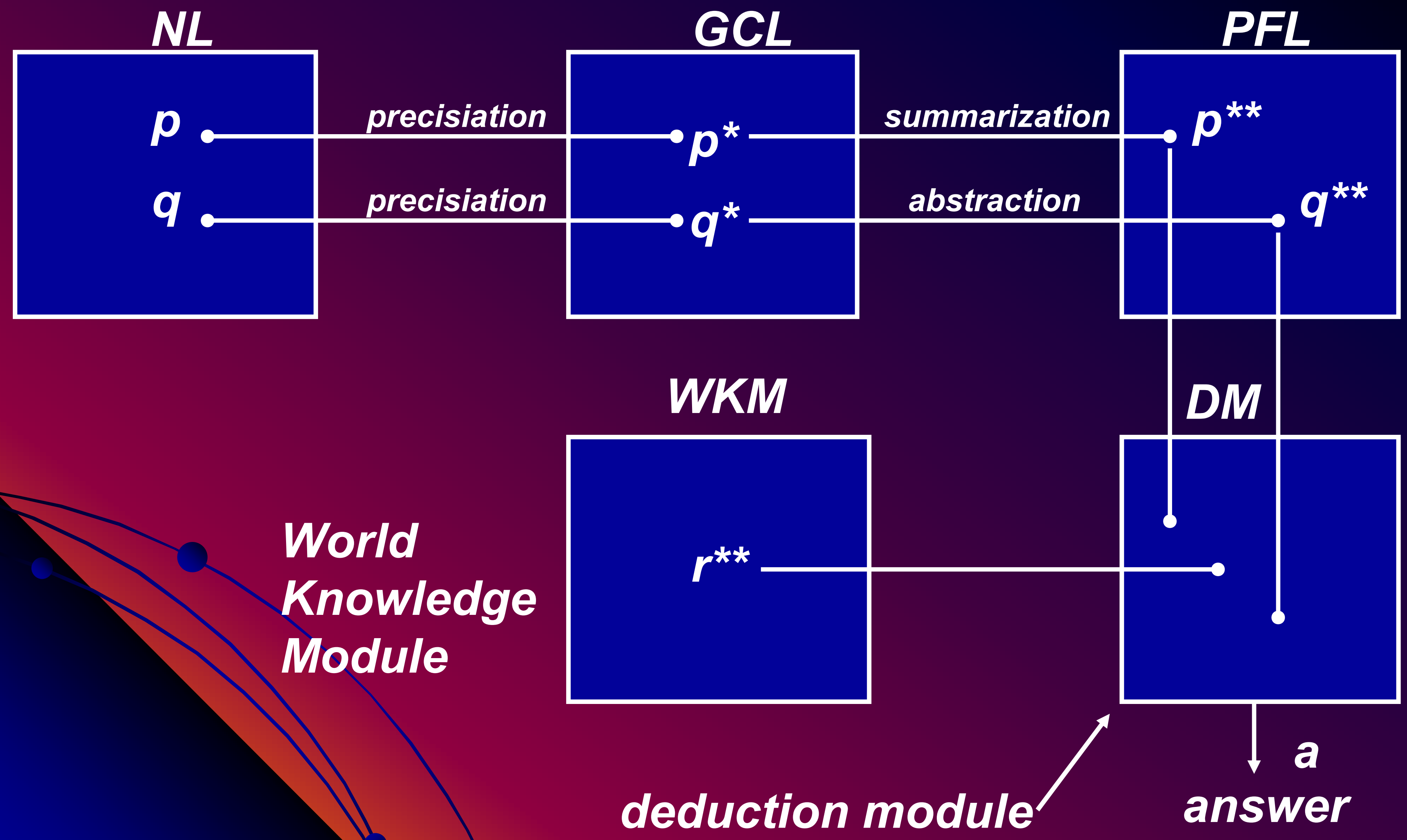
query: ?Name (A)|Desc (A)

query: Name (B)|Desc (B)

query: ?Attr (Name (A), Name (B))

PROTOFORMAL DEDUCTION

PROTOFORMAL DEDUCTION



PROTOFORMAL DEDUCTION

- Rules of deduction in the Deduction Database (DDB) are protoformal

  examples: (a) compositional rule of inference

symbolic

$$\frac{\begin{array}{c}X \text{ is } A \\ (X, Y) \text{ is } B\end{array}}{Y \text{ is } A \circ B}$$

$$\mu_B(v) = sup(\mu_A(u) \wedge \mu_B(u,v))$$

computational

(b) extension principle

$$\frac{\begin{array}{c}X \text{ is } A \\ Y = f(X)\end{array}}{Y = f(A)}$$

$$\mu_y(v) = sup_u(\mu_A(u))$$

Subject to: $v = f(u)$

symbolic

computational

RULES OF DEDUCTION

- ***Rules of deduction are basically rules governing generalized constraint propagation***
- ***The principal rule of deduction is the extension principle***

$$\frac{X \text{ is } A}{f(X,) \text{ is } B}$$

symbolic

$$\mu_B(v) = sup_u(\mu_A(u)$$

$$\text{Subject to: } v = f(u)$$

computational

GENERALIZATIONS OF THE EXTENSION PRINCIPLE

information = constraint on a variable

$$\frac{f(X) \text{ is } A}{g(X) \text{ is } B}$$

← *given information about X*

← *inferred information about X*

$$\mu_B(v) = \sup_u(\mu_A(f(u)))$$

Subject to: $v = g(u)$

CONTINUED

$$\frac{f(X_1, \ldots, X_n) \text{ is } A}{g(X_1, \ldots, X_n) \text{ is } B}$$

$$\mu_B(v) = \sup_u(\mu_A(f(u))$$

Subject to: $v = g(u)$

$(X_1, \ldots, X_n)$ is A

$\underline{g_j(X_1, \ldots, X_n) \text{ is } Y_j}$, $j=1, \ldots, n$

$(Y_1, \ldots, Y_n)$ is B

$$\mu_B(v) = \sup_u(\mu_A(f(u))$$

Subject to: $v = g(u)$

$j = 1, \ldots, n$

PROTOFORMAL DEDUCTION

Example:

most Swedes are tall $\longrightarrow$ *1/nΣCount(G[A] is R) is Q*

Height

PROTOFORMAL DEDUCTION RULE

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

$$\frac{\Sigma\mu_R(A_i) \text{ is } Q}{\Sigma\mu_S(A_i) \text{ is } T}$$

$$\mu_T(v) = sup_{A_1, \ldots, A_n}(\mu_Q(\Sigma_i \mu_R(A_i))$$

subject to

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

EXAMPLE OF DEDUCTION

p: Most Swedes are much taller than most Italians

q: What is the difference in the average height of Swedes and Italians?

PNL-based solution

Step 1. precisiation: translation of p into GCL

$S = \{S_1, \ldots, S_n\}$: population of Swedes

$I = \{I_1, \ldots, I_n\}$: population of Italians

g_i = height of S_i , $g = (g_1, \ldots, g_n)$

h_j = height of I_j , $h = (h_1, \ldots, h_n)$

$\mu_{ij} = \mu_{much.taller}(g_i, h_j)$ = degree to which S_i is much taller than I_j

CONTINUED

$r_i = \frac{1}{n}\Sigma_j \mu_{ij}$ = *Relative ΣCount of Italians in relation to whom S_i is much taller*

$t_i = \mu_{most}(r_i)$ = *degree to which S_i is much taller than most Italians*

$v = \frac{1}{m}\Sigma t_i$ = *Relative ΣCount of Swedes who are much taller than most Italians*

$ts(g, h) = \mu_{most}(v)$

$p \longrightarrow$ *generalized constraint on S and I*

$q: d = \frac{1}{m}\Sigma_i g_i - \frac{1}{n}\Sigma_j h_j$

CONTINUED

Step 2. Deduction via extension principle

$$\mu_q(d) = sup_{g,h}\, ts(g,h)$$

subject to

$$d = \frac{1}{m}\Sigma_i g_i - \frac{1}{n}\Sigma_j h_j$$

DEDUCTION PRINCIPLE

- *Point of departure: question, q*
- *Data: $D = (X_1/u_1, \ldots, X_n/u_n)$*

u_i *is a generic value of* X_i

- *Ans(q): answer to q*
- *If we knew the values of the* X_i*,* $u_1, \ldots, u_n$*, we could express Ans(q) as a function of the* u_i

$$Ans(q)=g(u_1, \ldots, u_n) \qquad u=(u_1, \ldots, u_n)$$

- *Our information about the* u_i*,* $I(u_1, \ldots, u_n)$ *is a generalized constraint on the* u_i*. The constraint is defined by its test-score function*

$$f(u)=f(u_1, \ldots, u_n)$$

CONTINUED

- Use the extension principle

$$\mu_{Ans(q)}(v) = sup_u(ts(u))$$

subject to

$$v = g(u)$$

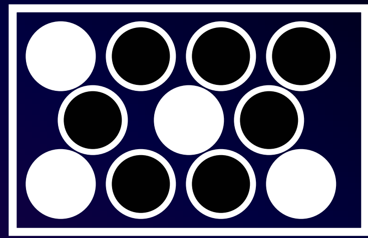
SUMMATION

- addition of significant question-answering capability to search engines is a complex, open-ended problem
- incremental progress, but not much more, is achievable through the use of bivalent-logic-based methods
- to achieve significant progress, it is imperative to develop and employ new methods based on computing with words, protoform theory, precisiated natural language and computational theory of precisiation of meaning
- The centerpiece of new methods is the concept of a generalized constraint

APPENDIX

DEDUCTION
THE BALLS-IN-BOX PROBLEM

Version 1. Measurement-based



A flat box contains a layer of black and white balls. You can see the balls and are allowed as much time as you need to count them

- q_1*: What is the number of white balls?*
- q_2*: What is the probability that a ball drawn at random is white?*
- q_1 *and* q_2 *remain the same in the next version*

DEDUCTION

Version 2. Perception-based

You are allowed n seconds to look at the box. n seconds is not enough to allow you to count the balls

You describe your perceptions in a natural language

p_1: there are about 20 balls

p_2: most are black

p_3: there are several times as many black balls as white balls

PT's solution?

MEASUREMENT-BASED

version 1

- a box contains 20 black and white balls
- over seventy percent are black
- there are three times as many black balls as white balls
- what is the number of white balls?
- what is the probability that a ball picked at random is white?

PERCEPTION-BASED

version 2

- a box contains about 20 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?

COMPUTATION (version 2)

- *measurement-based*

  X = number of black balls

  Y_2 number of white balls

  $X \geq 0.7 \bullet 20 = 14$

  $X + Y = 20$

  $X = 3Y$

  $X = 15$; $Y = 5$

  $p = 5/20 = .25$

- *perception-based*

  X = number of black balls

  Y = number of white balls

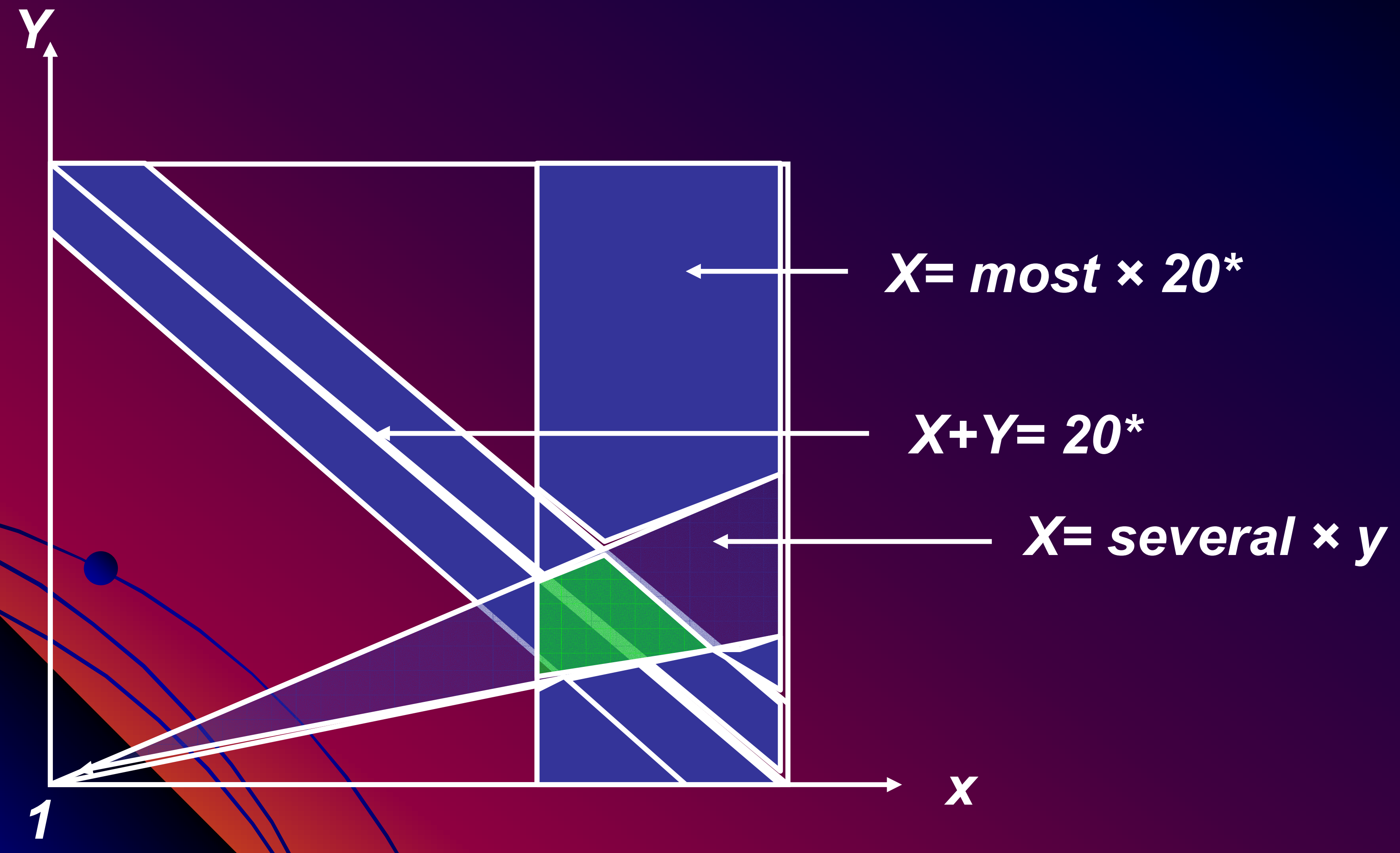
  $X = most \times 20^*$

  $X = several\ ^*Y$

  $X + Y = 20^*$

  $P = Y/N$

FUZZY INTEGER PROGRAMMING



RELEVANCE, REDUNDANCE AND DELETABILITY

DECISION TABLE

Name		A_1	A_j	A_n		D
$Name_1$		a_{11}	a_{1j}	a_{in}		d_1
.		.	.	.		.
$Name_k$		a_{k1}	a_{kj}	a_{kn}		d_1
$Name_{k+1}$		$a_{k+1, 1}$	$a_{k+1, j}$	$a_{k+1, n}$		d_2
.		.	.	.		.
$Name_l$		a_{l1}	a_{lj}	a_{ln}		d_l
.		.	.	.		.
$Name_n$		a_{m1}	a_{mj}	a_{mn}		d_r

A_j: j th symptom

a_{ij}: value of j th symptom of Name

D: diagnosis

REDUNDANCE ⟶ DELETABILITY

Name		A_1	A_j	A_n		D
.		.	.	.		.
$Name_r$		a_{r1}	*	a_{rn}		d_2
.		.	.	.		.

A_j is conditionally redundant for $Name_r$, A, is a_{r1}, A_n is a_{rn}
*If D is d_s for all possible values of A_j in **

A_j is redundant if it is conditionally redundant for all values of Name

- *compactification algorithm (Zadeh, 1976); Quine-McCluskey algorithm*

RELEVANCE

D is $?d$ if A_j is a_{rj}

constraint on A_j induces a constraint on D

example: (blood pressure is high) constrains D

(A_j is a_{rj}) is uniformative if D is unconstrained

A_j is irrelevant if it A_j is uniformative for all a_{rj}

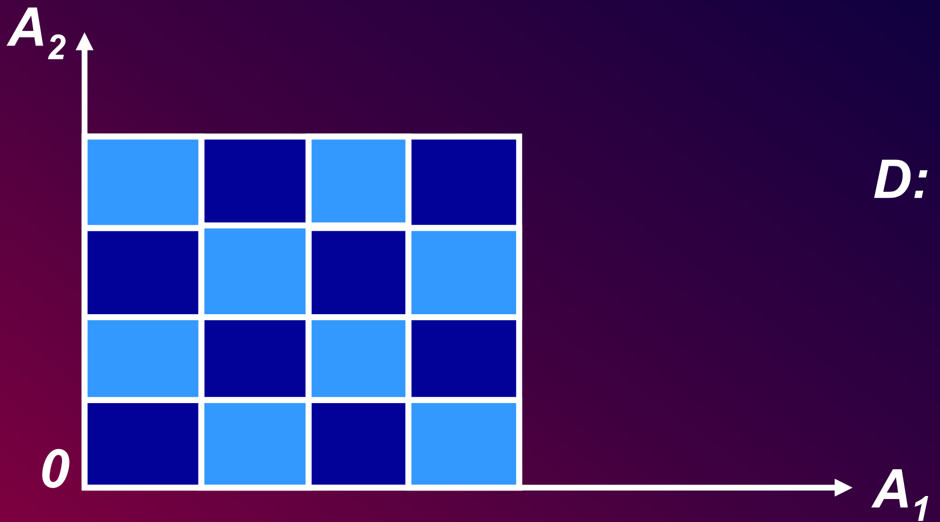
irrelevance $\not\longrightarrow$ deletability

IRRELEVANCE (UNINFORMATIVENESS)

Name	A_1	A_j	A_n		D
$Name_r$	.	a_{ij}	.		d_1 . d_1
$Name_{i+s}$	.	a_{ij}	.		d_2 . d_2

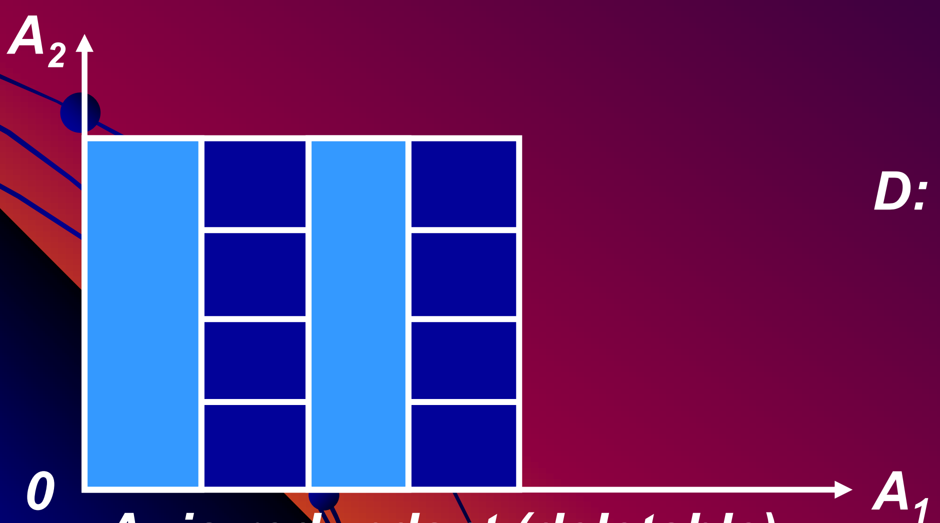
(A_j is a_{ij}) is irrelevant (uninformative)

EXAMPLE



D: black or white

A_1 and A_2 are irrelevant (uninformative) but not deletable



D: black or white

A_2 is redundant (deletable)

KEY POINT—THE ROLE OF FUZZY LOGIC

- *Existing approaches to the enhancement of web intelligence are based on classical, Aristotelian, bivalent logic and bivalent-logic-based probability theory. In our approach, bivalence is abandoned. What is employed instead is fuzzy logic—a logical system which subsumes bivalent logic as a special case.*
- *Fuzzy logic is not fuzzy*
- *Fuzzy logic is a precise logic of fuzziness and imprecision*
- *The centerpiece of fuzzy logic is the concept of a generalized constraint.*

- *In bivalent logic, BL, truth is bivalent, implying that every proposition, p, is either true or false, with no degrees of truth allowed*

- *In multivalent logic, ML, truth is a matter of degree*

- *In fuzzy logic, FL:*
  - *everything is, or is allowed to be, to be partial, i.e., a matter of degree*
  - *everything is, or is allowed to be, imprecise (approximate)*
  - *everything is, or is allowed to be, granular (linguistic)*
  - *everything is, or is allowed to be, perception based*

CONTINUED

- ***The generality of fuzzy logic is needed to cope with the great complexity of problems related to search and question-answering in the context of world knowledge; to deal computationally with perception-based information and natural languages; and to provide a foundation for management of uncertainty and decision analysis in realistic settings***

Factual Information About the Impact of Fuzzy Logic

PATENTS

- Number of fuzzy-logic-related patents applied for in Japan: 17,740
- Number of fuzzy-logic-related patents issued in Japan: 4,801
- Number of fuzzy-logic-related patents issued in the US: around 1,700

Count of papers containing the word "fuzzy" in title, as cited in INSPEC and MATH.SCI.NET databases.

Compiled by Camille Wanat, Head, Engineering Library, UC Berkeley,

December 22, 2004

Number of papers in INSPEC and MathSciNet which have "fuzzy" in their titles:

INSPEC - "fuzzy" in the title

1970-1979: 569

1980-1989: 2,404

1990-1999: 23,207

2000-present: 14,172

Total: 40,352

MathSciNet - "fuzzy" in the title

1970-1979: 443

1980-1989: 2,465

1990-1999: 5,483

2000-present: 3,960

Total: 12,351

JOURNALS (“fuzzy” or “soft computing” in title)

1. *Fuzzy Sets and Systems*
2. *IEEE Transactions on Fuzzy Systems*
3. *Fuzzy Optimization and Decision Making*
4. *Journal of Intelligent & Fuzzy Systems*
5. *Fuzzy Economic Review*
6. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*
7. *Journal of Japan Society for Fuzzy Theory and Systems*
8. *International Journal of Fuzzy Systems*
9. *Soft Computing*
10. *International Journal of Approximate Reasoning--Soft Computing in Recognition and Search*
11. *Intelligent Automation and Soft Computing*
12. *Journal of Multiple-Valued Logic and Soft Computing*
13. *Mathware and Soft Computing*
14. *Biomedical Soft Computing and Human Sciences*
15. *Applied Soft Computing*

APPLICATIONS

The range of application-areas of fuzzy logic is too wide for exhaustive listing. Following is a partial list of existing application-areas in which there is a record of substantial activity.

1. Industrial control
2. Quality control
3. Elevator control and scheduling
4. Train control
5. Traffic control
6. Loading crane control
7. Reactor control
8. Automobile transmissions
9. Automobile climate control
10. Automobile body painting control
11. Automobile engine control
12. Paper manufacturing
13. Steel manufacturing
14. Power distribution control
15. Software engineerinf
16. Expert systems
17. Operation research
18. Decision analysis
19. Financial engineering
20. Assessment of credit-worthiness
21. Fraud detection
22. Mine detection
23. Pattern classification
24. Oil exploration
25. Geology
26. Civil Engineering
27. Chemistry
28. Mathematics
29. Medicine
30. Biomedical instrumentation
31. Health-care products
32. Economics
33. Social Sciences
34. Internet
35. Library and Information Science

Product Information Addendum 1

This addendum relates to information about products which employ fuzzy logic singly or in combination. The information which is presented came from SIEMENS and OMRON. It is fragmentary and far from complete. Such addenda will be sent to the Group from time to time.

SIEMENS:

* washing machines, 2 million units sold
* fuzzy guidance for navigation systems (Opel, Porsche)
* OCS: Occupant Classification System (to determine, if a place in a car is occupied by a person or something else; to control the airbag as well as the intensity of the airbag). Here FL is used in the product as well as in the design process (optimization of parameters).
* fuzzy automobile transmission (Porsche, Peugeot, Hyundai)

OMRON:

* fuzzy logic blood pressure meter, 7.4 million units sold, approximate retail value $740 million dollars

Note: If you have any information about products and or manufacturing which may be of relevance please communicate it to Dr. Vesa Niskanen vesa.a.niskanen@helsinki.fi and Masoud Nikravesh Nikravesh@cs.berkeley.edu .

Product Information Addendum 2

This addendum relates to information about products which employ fuzzy logic singly or in combination. The information which is presented came from Professor Hideyuki Takagi, Kyushu University, Fukuoka, Japan. Professor Takagi is the co-inventor of neurofuzzy systems. Such addenda will be sent to the Group from time to time.

Facts on FL-based systems in Japan (as of 2/06/2004)

1. Sony's FL camcorders

Total amount of camcorder production of all companies in 1995-1998 times Sony's market share is the following. Fuzzy logic is used in all Sony's camcorders at least in these four years, i.e. total production of Sony's FL-based camcorders is 2.4 millions products in these four years.

1,228K units X 49% in 1995
1,315K units X 52% in 1996
1,381K units X 50% in 1997
1,416K units X 51% in 1998

2. FL control at Idemitsu oil factories

Fuzzy logic control is running at more than 10 places at 4 oil factories of Idemitsu Kosan Co. Ltd including not only pure FL control but also the combination of FL and conventional control.

They estimate that the effect of their FL control is more than 200 million YEN per year and it saves more than 4,000 hours per year.

3. Canon

Canon used (uses) FL in their cameras, camcorders, copy machine, and stepper alignment equipment for semiconductor production. But, they have a rule not to announce their production and sales data to public.

Canon holds 31 and 31 established FL patents in Japan and US, respectively.

4. Minolta cameras

Minolta has a rule not to announce their production and sales data to public, too.

whose name in US market was Maxxum 7xi. It used six FL systems in a camera and was put on the market in 1991 with 98,000 YEN (body price without lenses). It was produced 30,000 per month in 1991. Its sister cameras, alpha-9xi, alpha-5xi, and their successors used FL systems, too. But, total number of production is confidential.

5. FL plant controllers of Yamatake Corporation

Yamatake-Honeywell (Yamatake's former name) put FUZZICS, fuzzy software package for plant operation, on the market in 1992. It has been used at the plants of oil, oil chemical, chemical, pulp, and other industries where it is hard for conventional PID controllers to describe the plan process for these more than 10 years.

They planed to sell the FUZZICS 20 - 30 per year and total 200 million YEN.

As this software runs on Yamatake's own control systems, the software package itself is not expensive comparative to the hardware control systems.

6. Others

Names of 225 FL systems and products picked up from news articles in 1987 - 1996 are listed at http://www.adwin.com/elec/fuzzy/note_10.html in Japanese.)

Note: If you have any information about products and or manufacturing which may be of relevance please communicate it to Dr. Vesa Niskanen vesa.a.niskanen@helsinki.fi and Masoud Nikravesh Nikravesh@cs.berkeley.edu , with cc to me.