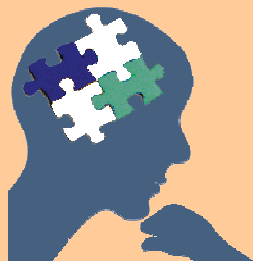# Resolving the Accuracy-Informativeness Conflict in Question Answering:
A New Grain-Control Model

Rakefet Ackerman and Morris Goldsmith
University of Haifa, Haifa, Israel

מעמק"ה – המכון
לעיבוד מידע
וקבלת החלטות
IIPDM - Institute of
Information Processing
and Decision Making

אוניברסיטת חיפה
University of Haifa

Question Answering Under Uncertainty

"What time did the incident occur?"

Control over Grain-Size   **Confidence**

"5:42PM"   **Best guess**   (10%)

Late afternoon   **Highly likely** (90%)
("5:00 – 6:30PM")

Accuracy – Informativeness Trade-off

(Yaniv and Foster, 1995, 1997)

Accuracy – Informativeness Trade-off

Confidence | Informativeness

100% — "5:00-6:30" — Low

90%

Correctness preference

"5:15-6:00"

70%

20%

Informativeness preference

High

"5:42"

10%

Interval width of Candidate Answer

3
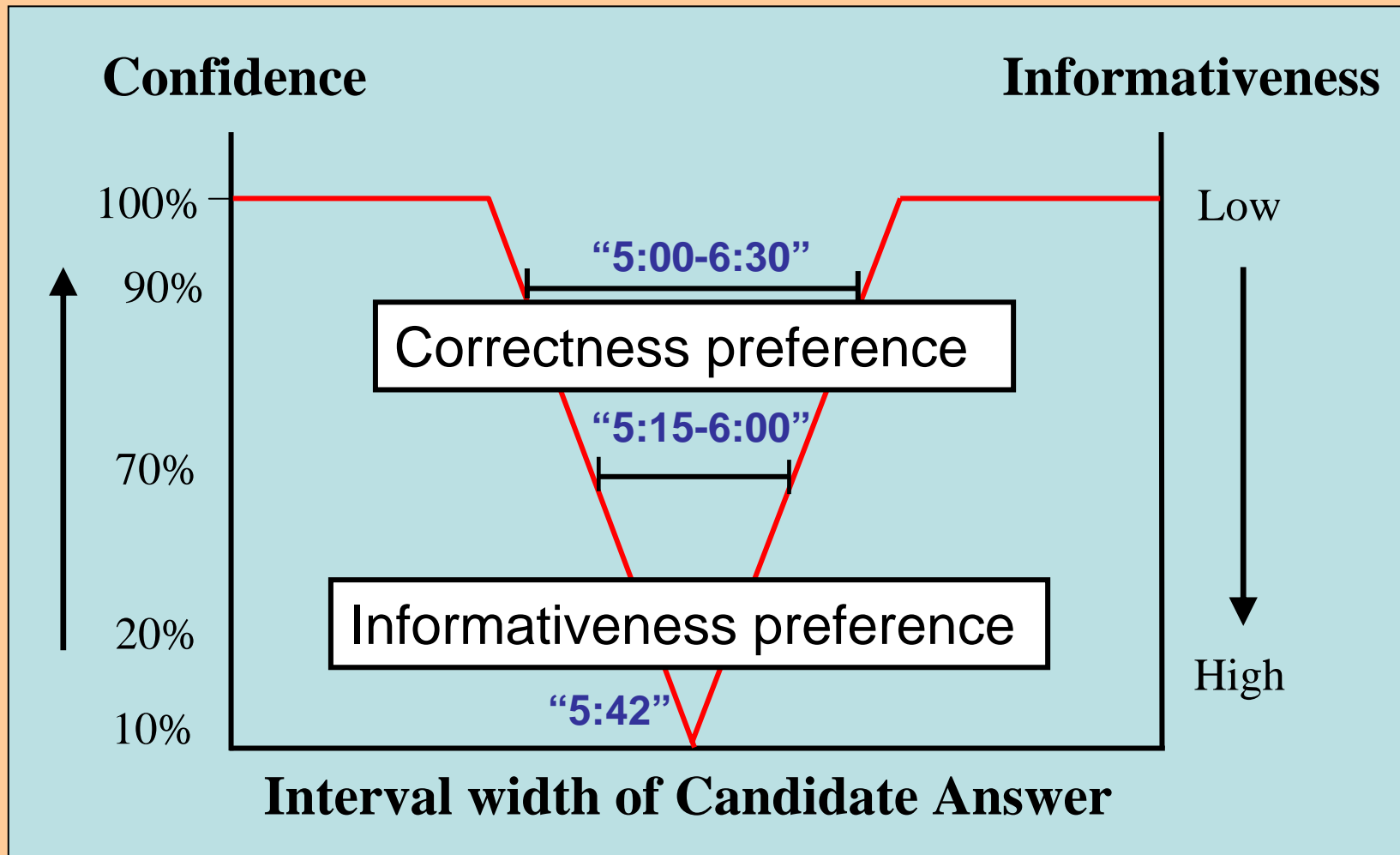
Research Overview

Which aim is preferred:
Correctness or Informativeness?

Answering process models:

- Starting point: **Satisficing Model**
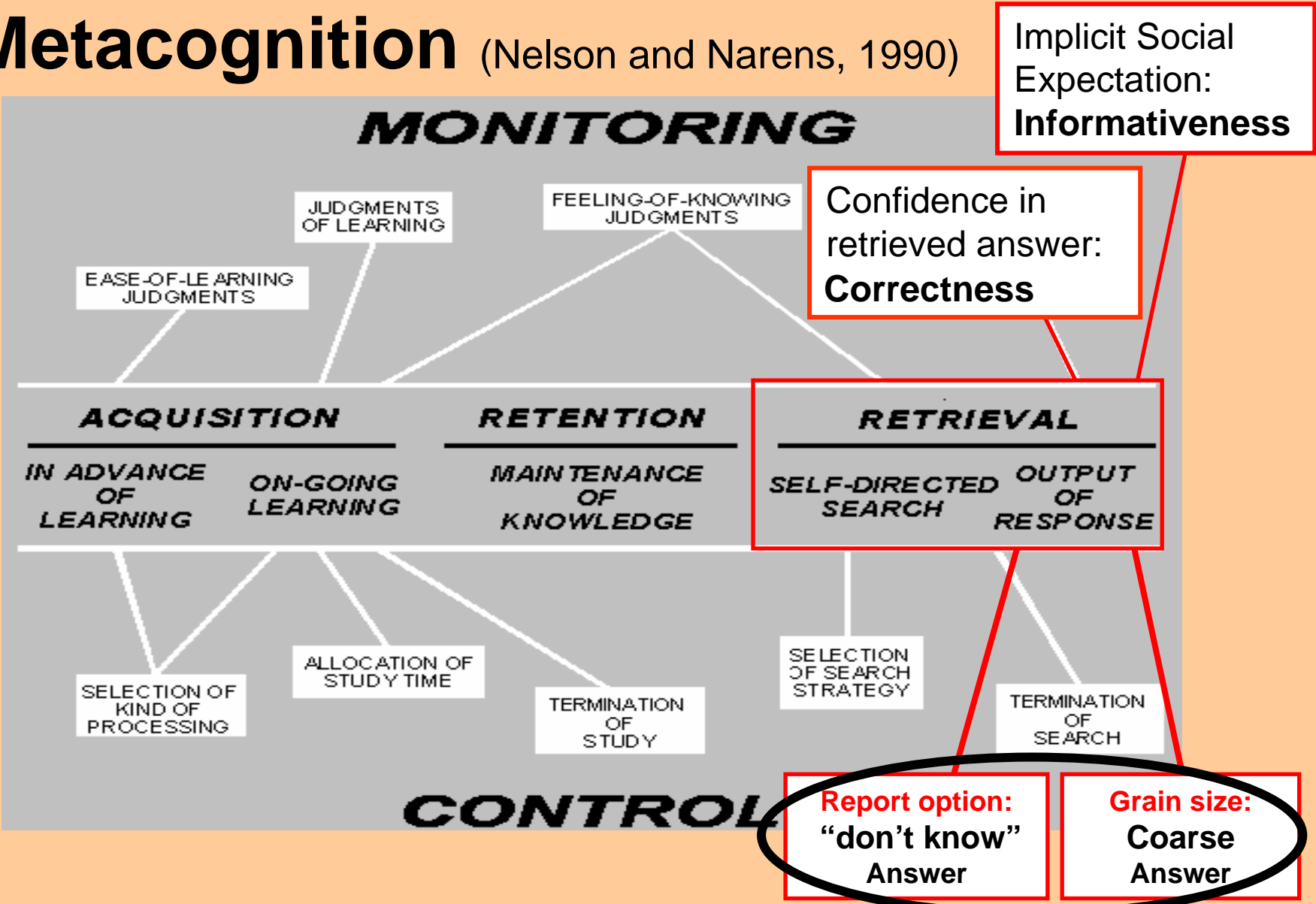
  (Goldsmith, Koriat, & Weinberg-Eliezer, 2002)

  - Adequate for high knowledge situations

- Extended model: **Dual-Criterion Model**

  - Adds low knowledge answering process
  - Use of "don't know" answer

4

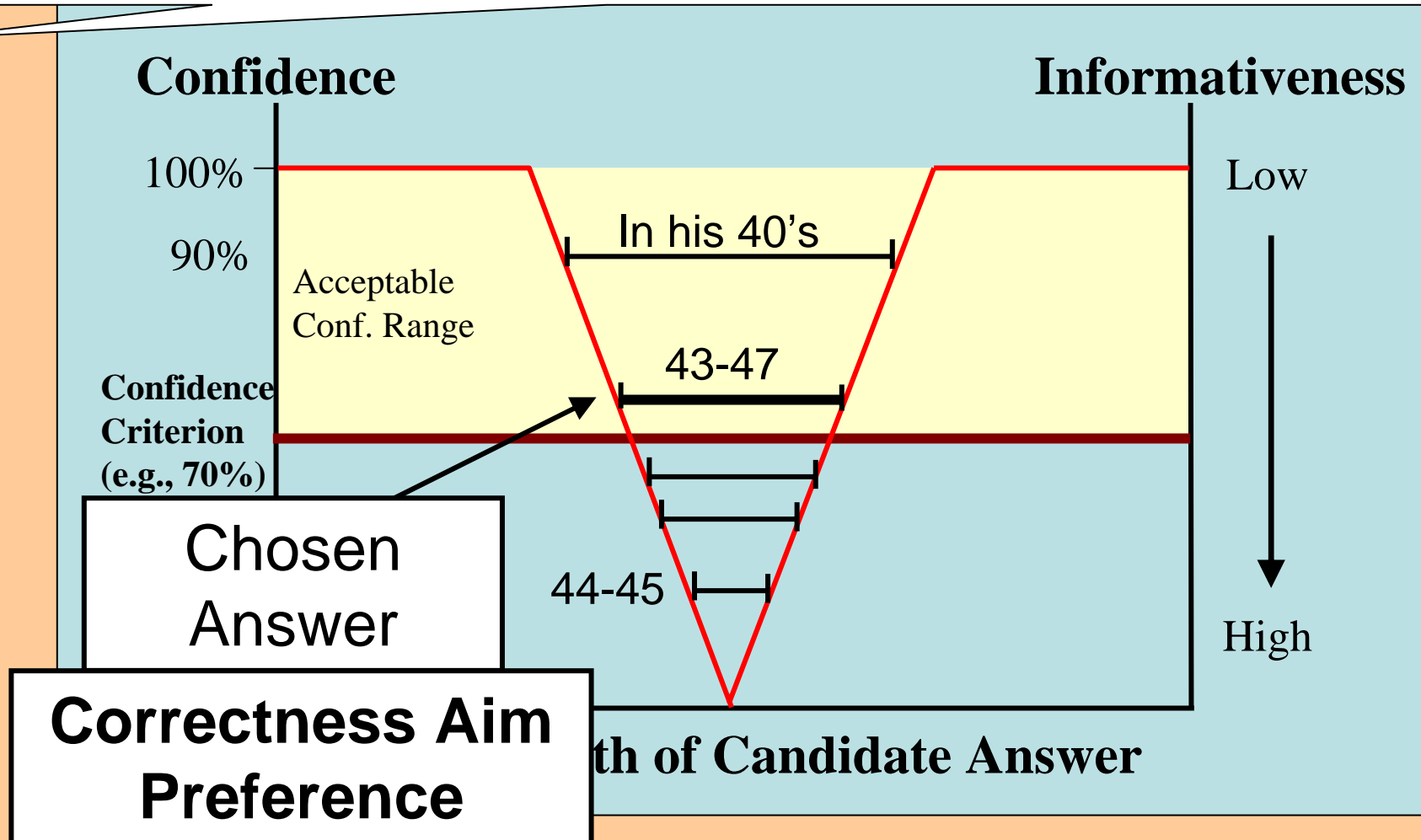# **Metacognition** (Nelson and Narens, 1990)

Koriat & Goldsmith (1996); Goldsmith & Koriat (1999)

5

Satisficing Model

(Goldsmith, Koriat, and Weinberg-Eliezer, 2002)

Example: How old was Kennedy at the time of his assassination?
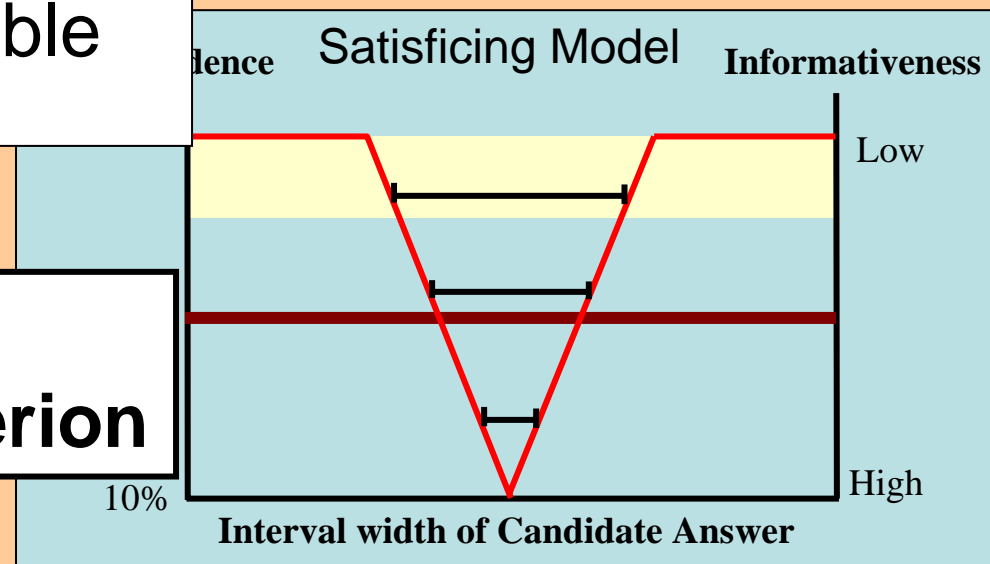
The Current Research

What is the answering process in case of <u>low knowledge</u>?

🌐 Ridiculously coarse answers

But…

Ridiculously coarse answers are not socially acceptable (Grice, 1975)

Minimum Informativeness Criterion

Satisficing Model

...dence | Informativeness

Low

High

10%

Interval width of Candidate Answer

The Current Research

What is the answering process in case of <u>low knowledge</u>?

🌐 Ridiculously coarse answers

What else?

🌐 Low confidence answers

🌐 "Don't know"

⬇

Contrary to the Satisfying Model

Satisfying Model

Confidence   **Informativeness**

Low

Conf. Criterion 70%

10%   **Interval width of Candidate Answer**   High

Dual-Criterion Model
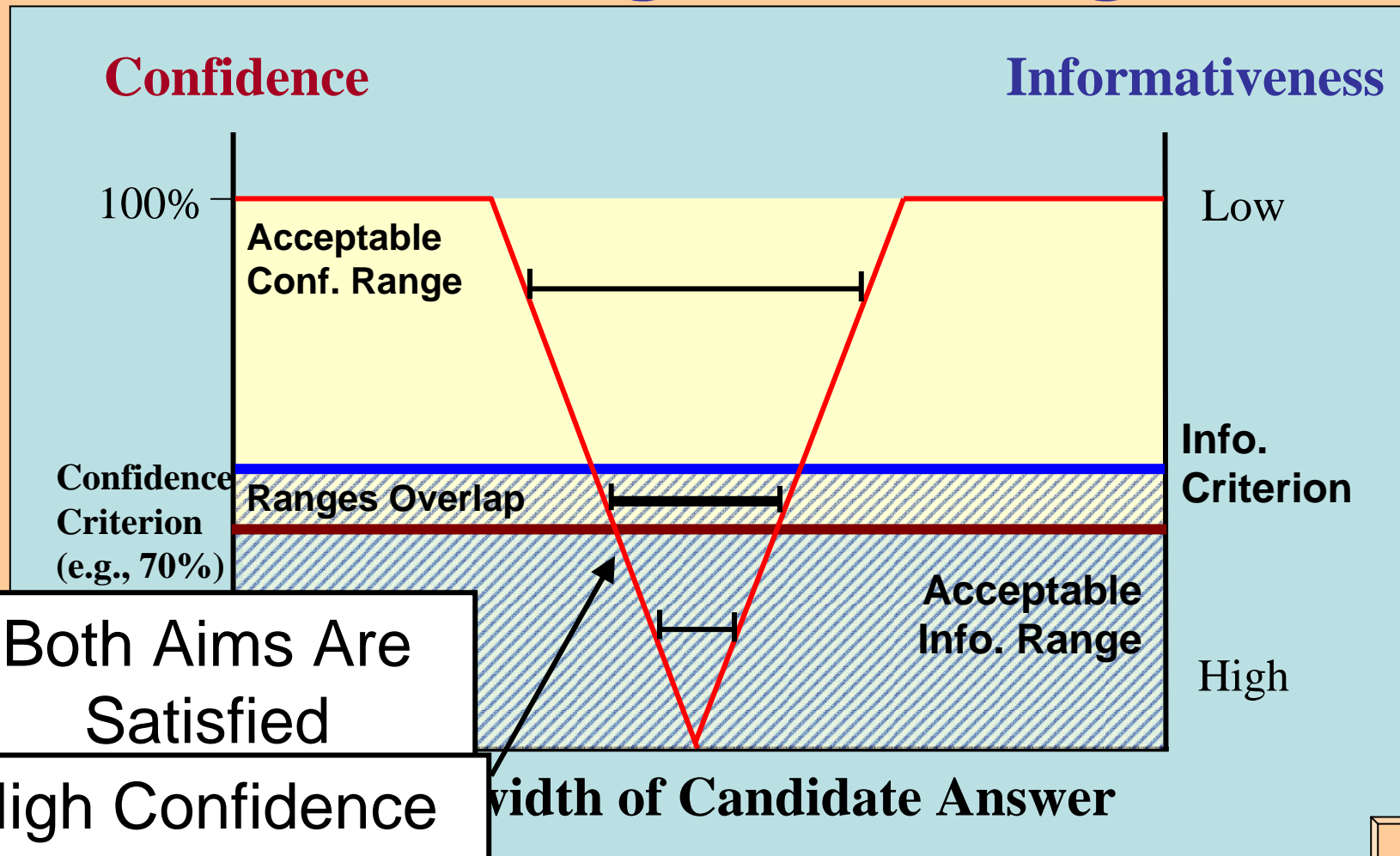
Satisfiing <u>Model</u>:

- Minimum Confidence Criterion
    - Maximum Informativeness

<u>Extensions</u>:

- Minimum Informativeness Criterion
- Two knowledge states:
    - **Satisficing Knowledge**
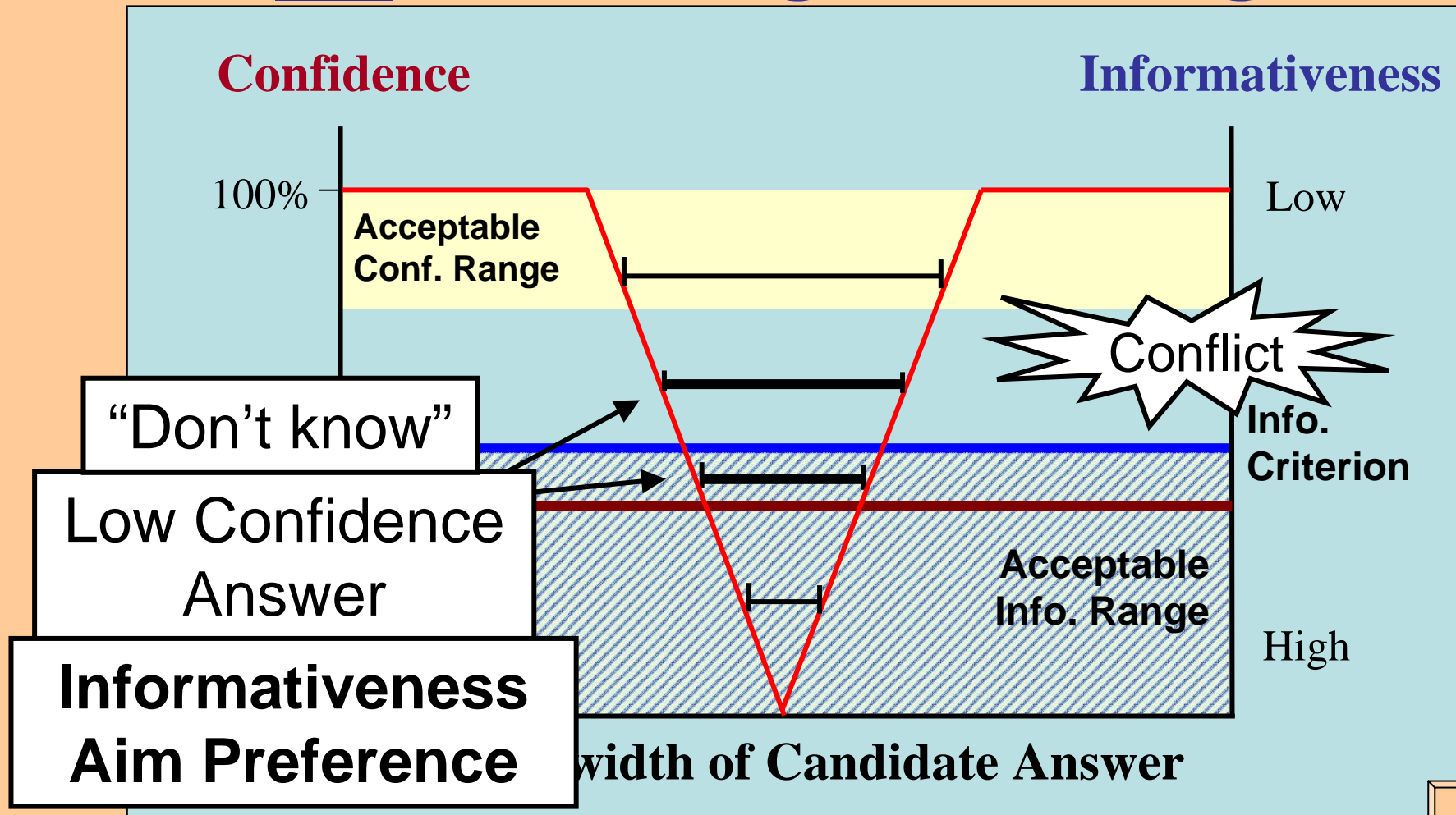   - **Unsatisficing Knowledge**

9
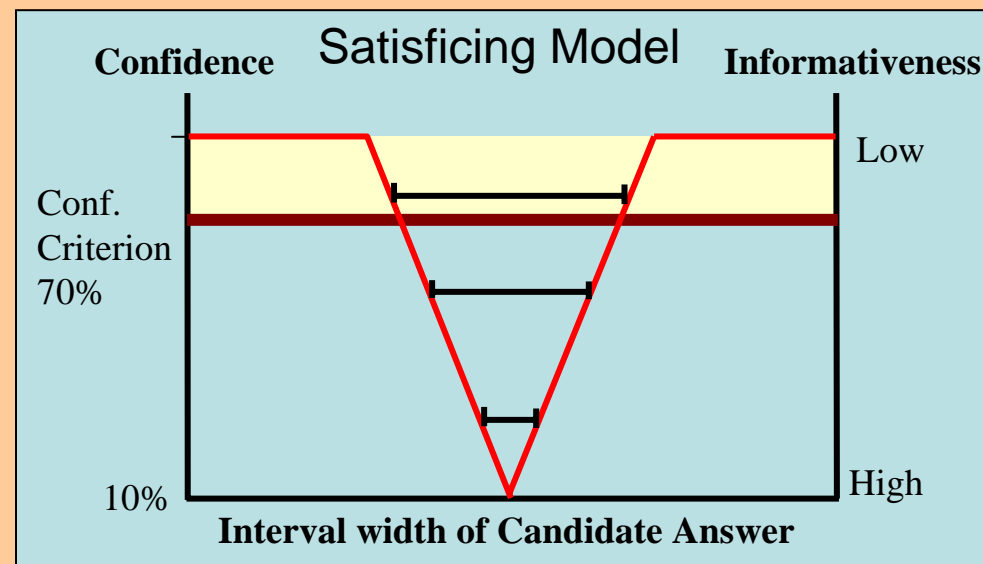
Dual-Criterion Model
Satisficing Knowledge

Confidence

Informativeness

100%

Low

Acceptable Conf. Range

Info. Criterion

Confidence Criterion (e.g., 70%)

Ranges Overlap

Acceptable Info. Range

High

Both Aims Are Satisfied

High Confidence Answer

Width of Candidate Answer

Dual-Criterion Model
<u>Un</u>satisficing Knowledge

Confidence

Informativeness

100% —

Low

Acceptable Conf. Range

Conflict

"Don't know"

Info. Criterion

Low Confidence Answer

Acceptable Info. Range

High

Informativeness Aim Preference

width of Candidate Answer

Experiments

Goal: Analyze the answering process in case of <u>low knowledge</u>:

Minimum Confidence Criterion	**+**	Minimum Informativeness Criterion	**=**

Low Confidence Answers
Experiment 1

"Don't know" Answers
Experiment 2
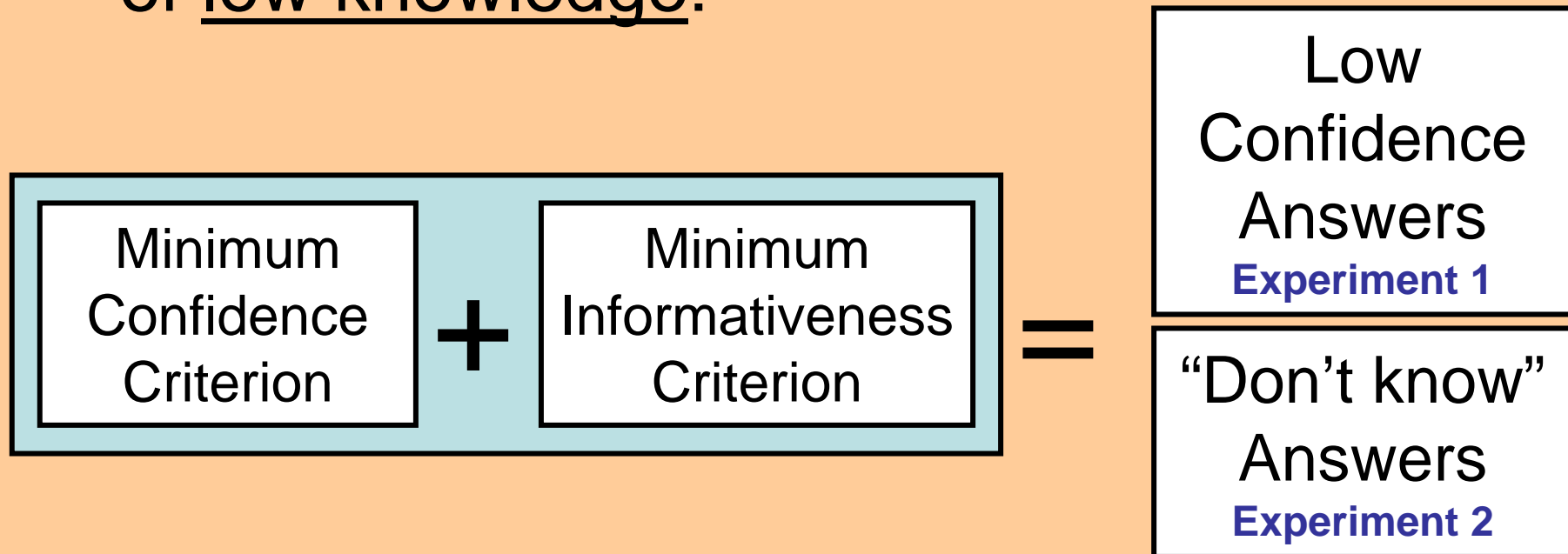
Experiment 1

- Materials

  Numerical general knowledge questions:

    20 medium difficulty – Moderate Knowledge (MK)
    20 very-hard   – Low Knowledge (LK)

- Two phase procedure:

    Free-grain (from ___ to ___) + Confidence rating
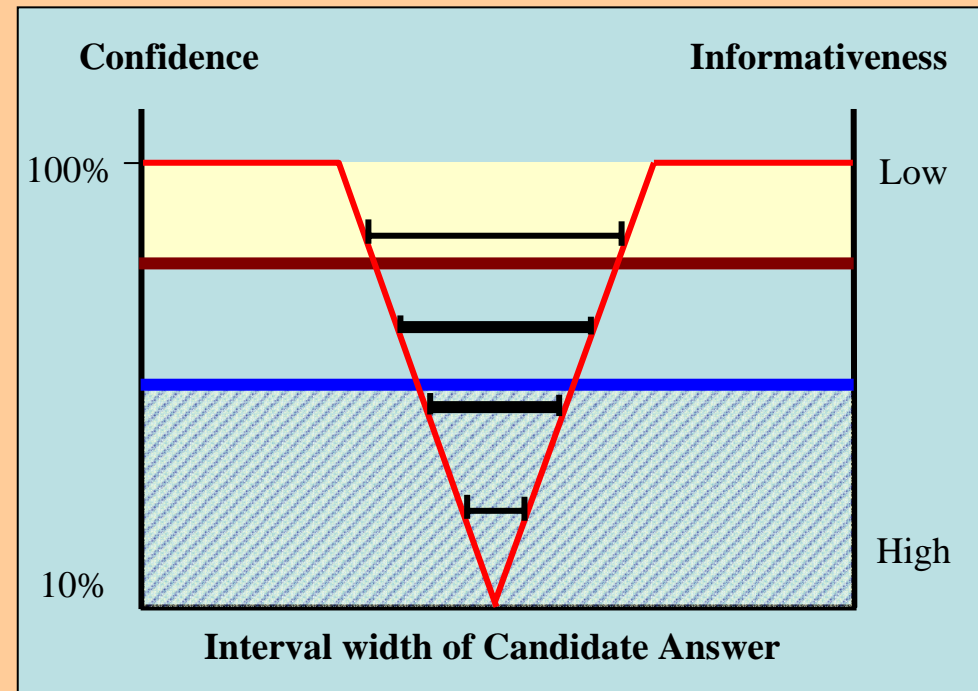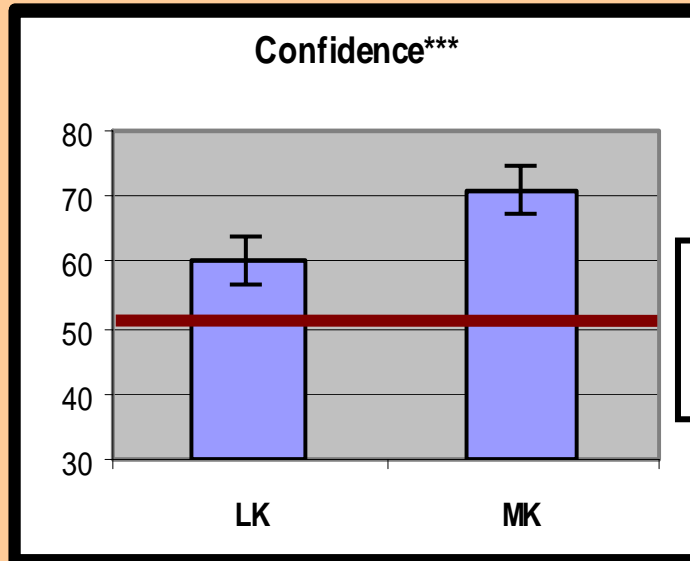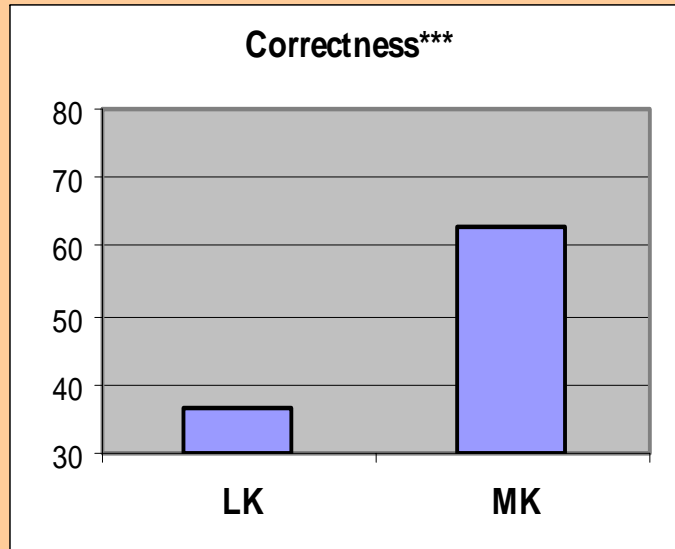   Fixed-grain (e.g. 10 countries interval)
                                                    + Confidence rating

- Predictions:

  1. Substantial amount of confidence criterion violations

  2. More violations for LK than for MK questions

Experiment 1 - Results

Correctness***

Confidence***

Confidence   Informativeness

Interval width of Candidate Answer

**Contrary the
Satisficing Model?**

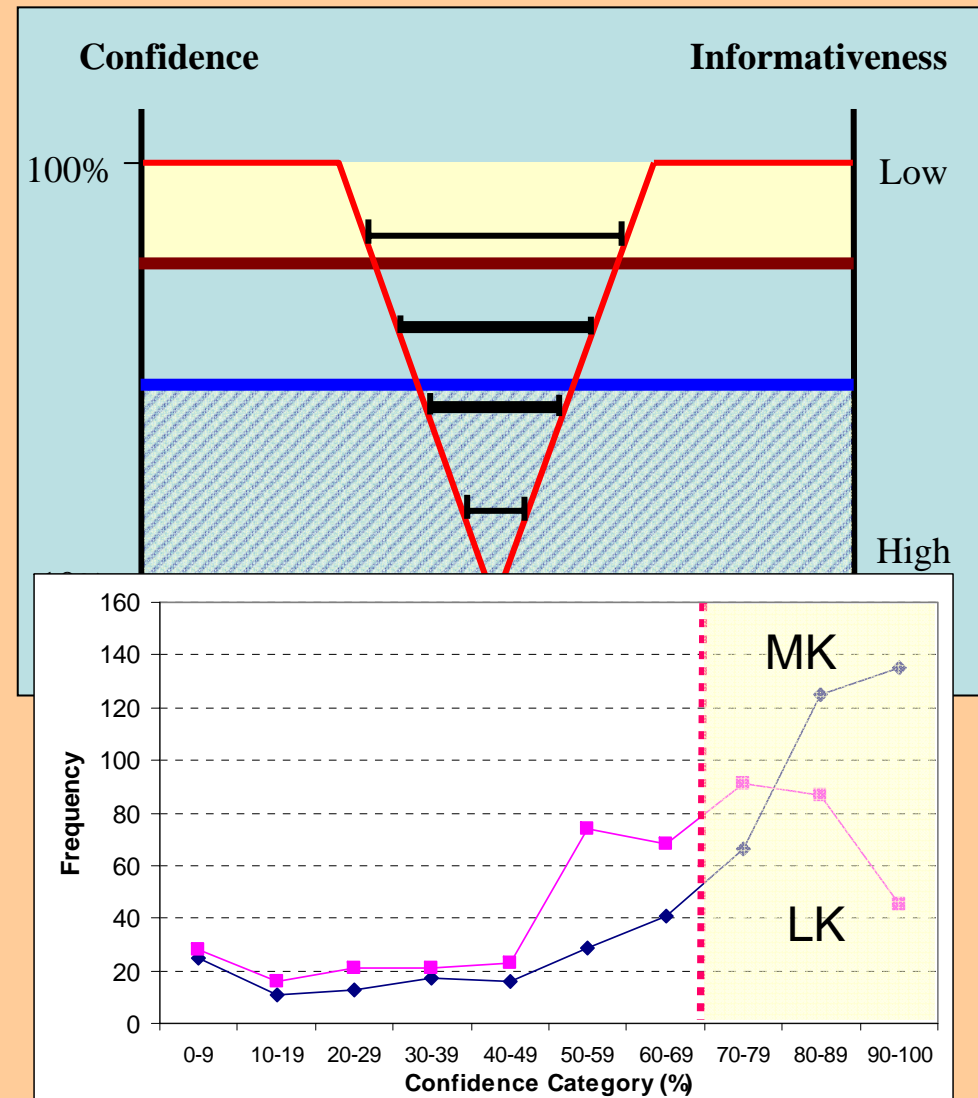*** p<.001

Experiment 1 – Results
Violations from confidence criterion

<u>Main findings:</u>

1. Substantial amount of confidence criterion violations

2. More for Low Knowledge (LK) than for Moderate Knowledge (MK)

Informativeness given priority over correctness

Contrary to the Satisficing Model

Experiment 2

- Materials

  Numerical general knowledge questions:

  20 medium difficulty – Moderate Knowledge (MK)

  20 very-hard   – Low Knowledge (LK)

- Two phase procedure:

  Free-grain + Confidence rating or **"DON'T KNOW"**
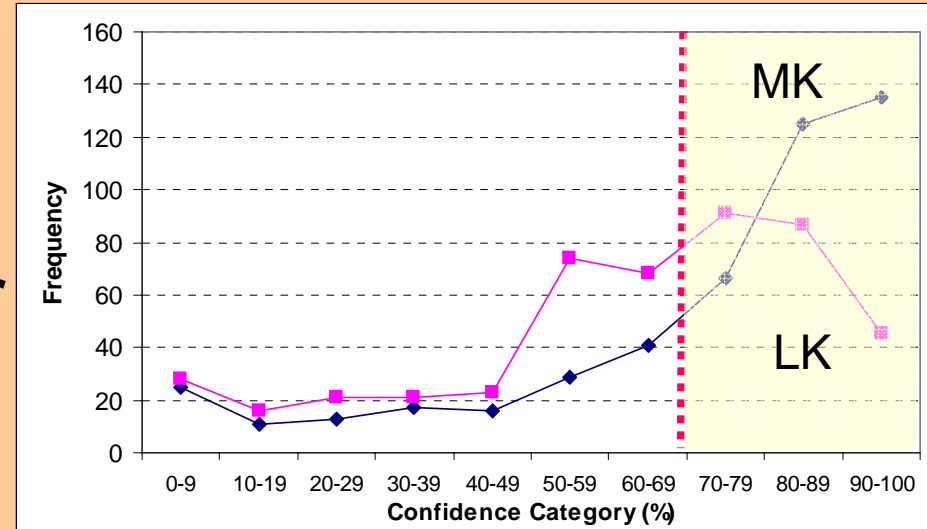
  Fixed-grain + Confidence rating

- Predictions:

  1. Reduced use of low confidence levels

  2. Reduced difference between question types in Confidence Criterion violations

Experiment 1+2 – Results
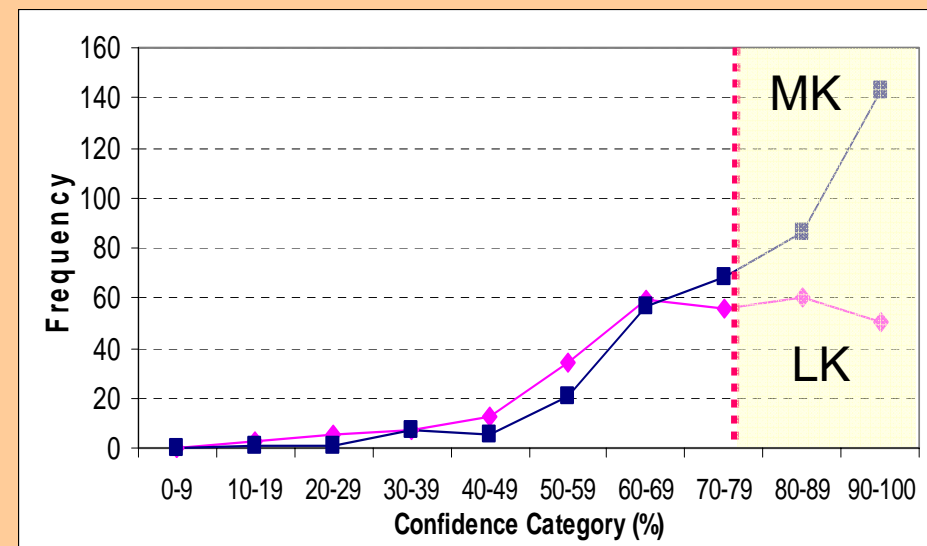## Violations from confidence criterion

Experiment 1:
More criterion violations for Low Knowledge (LK) than for Moderate Knowledge (MK)

Experiment 2:
Answerers avoid criterion violations by "don't know"

"don't know" as Conflict resolution

Summary

Research question:

> Which aim is preferred:
> **Correctness or Informativeness**?

Dual-Criterion Model:

- Minimum Confidence **+ Informativeness** Criteria
- Two knowledge states:
  - Satisficing Knowledge – Both aims are satisfied
  - **Unsatisficing Knowledge** – Conflict

| "Don't know" | Low Confidence Answer |

Real-life Consequences

- What do you mean by "I don't know"?

    "My knowledge is insufficient to support an answer that is both reasonably informative and reasonably likely to be correct."

- Witness testimony in low knowledge situations:

    "The whole truth and nothing but the truth"

    Witnesses should be directed explicitly which criterion to prefer

Rakefet Ackerman   **rakefet@research.haifa.ac.il**