

CMPT 880/890

Evaluating solutions

MOTD

- “Computer science meets every criterion for being a science, but it has a self-inflicted credibility problem.”
- “50% of [CS papers] proposing models or hypotheses *did not test them*. In other fields of science this fraction was about 10%...Our failure to test allows many unsound ideas to be tried in practice and lowers the credibility of our field as a science.”
 - (Denning)

Outline

- What is an evaluation?
- Evaluation methods
- Types of studies
- Validity

What is an evaluation?

- A traditional problem for CS students
- A way to determine whether your solution is a *good* solution to the problem you were trying to solve

An example

- You have developed a graphics technique for rendering crayon drawings.
- How should you evaluate your technique?

Issues for evaluators

- What problem are you trying to solve?
- Does your solution actually solve the problem?
 - ...or does it solve a different problem?
 - if they don't match, you can't design a good evaluation
- What bad thing results from the problem?
 - remember your motivation
 - why the problem was worth solving

Example

- Problem: people get lost in virtual worlds
- So what: extra time to navigate; frustration
- Solution: overview maps
- Evaluation:

Example

- Problem: designers don't know how teenagers use mobile phones
- So what: lack of knowledge → poor designs → less \$
- Solution: observe teenagers using mobile phones
- Evaluation:

Example

- Problem: Multi-tap on mobile phones is slow
- So what: too much time spent on SMS
- Solution: Tilt text
- Evaluation:

Appropriate evaluation methods

- There are many ways that you can gather evidence to show that your solution is a good one
- The most important question is whether your methods are appropriate for what you want to achieve
- What kind of a contribution are you making?
- What is the appropriate type of evaluation?

Evaluation methods

- Field studies
- Surveys and interviews
- Laboratory studies with people
- Simulations (laboratory studies without people)
- Proofs and demonstrations
- Analysis and argument

Evaluation by demonstration

- If your contribution is to be the very first to solve a problem, you only have to demonstrate that you have solved it
- For example, if your main claim is “I have built a robot duck,” then all you have to do is show the duck

Evaluation by argumentation

- Evaluations do not always require empirical evidence
- You only need to ensure that your readers accept your main claim
- Can you convince them through analysis?
- For example, an approach that is clearly better:
 - why is it clearly better?
 - how can you make the difference clear?
 - e.g., count steps that are done in the algorithm
- Studies or simulations are not needed if the difference is obvious

Some terminology

- Qualitative vs. quantitative
- Formative vs. summative

Some terminology

- Quantitative
 - results in numerical data
- Qualitative
 - results in descriptive textual data
- Formative
 - provides information at the start of the research
 - exploratory
- Summative
 - provides information at the end of the research
 - 'stamp of approval'

McGrath: methodology matters

Research Methods: Opportunities and Limitations

- Methods enable but also limit evidence.
- All methods are valuable, but all have weaknesses or limitations.
- You can offset the different weaknesses of various methods by using multiple methods.
- You can choose such multiple methods so that they have *patterned diversity*; that is so that strengths of some methods offset weaknesses of others.

Research Methods: Opportunities and Limitations

- “Credible empirical knowledge requires convergence of evidence across studies based on different methods.”

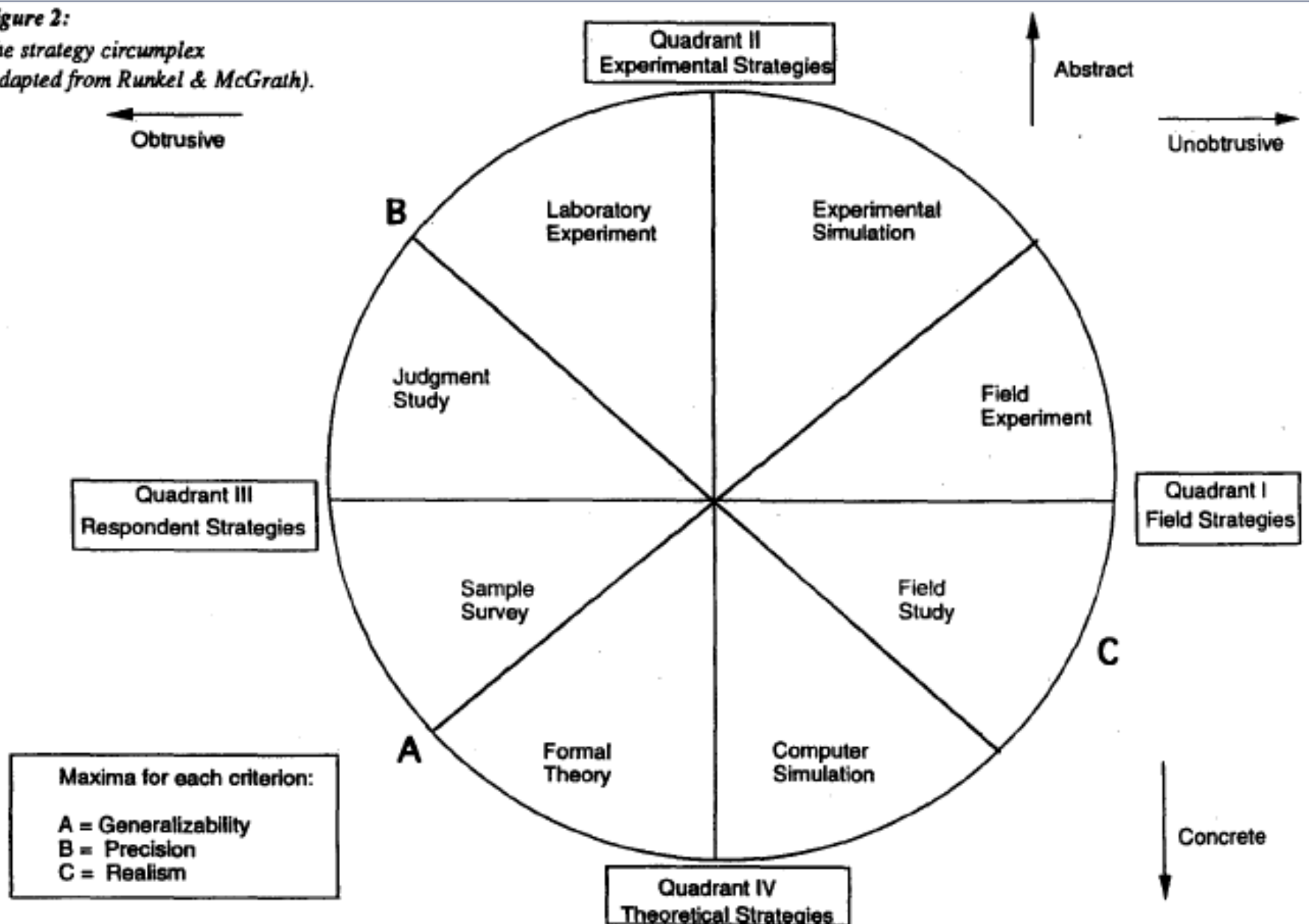
Research Strategies

- When you gather research evidence, you are always trying to maximize three desirable features:
- **Generalizability** of the evidence
- **Precision** of the measurement of what is being studied (and precision of control over extraneous factors that are not being studied)
- **Realism** of the situation or context within which the evidence is gathered

Research Strategies

- Although you always want to maximize A, B, and C simultaneously, you cannot do so.

Figure 2:
The strategy circumplex
(adapted from Runkel & McGrath).



Quadrant I: The Field Strategies

Quadrant I: The Field Strategies

- Field study
 - the researcher sets out to make direct observations of “natural,” ongoing systems, while disturbing those systems as little as possible.
- Field experiment
 - also works within an ongoing natural system as unobtrusively as possible, except for intruding on that system by manipulating one major feature of that system.

Quadrant I: The Field Strategies

- The behavior system under study is “natural:”
 - it would occur whether or not the researcher were there and whether or not it were being observed

Quadrant II: The Experimental Strategies

Quadrant II: The Experimental Strategies

- Laboratory experiment
 - the investigator deliberately concocts a situation, defines the rules for its operation, and then induces individuals to enter the concocted system and engage in the behaviors called for by its rules and circumstances.
- Experimental simulation
 - the researcher attempts to achieve much of the precision and control of the laboratory experiment but to gain some of the realism of field studies.

Quadrant II: The Experimental Strategies

- The laboratory experiment and the experimental simulation are strategies that involve systems that would not exist at all were it not for the researcher's interest in doing the study.

Quadrant III: The Respondent Strategies

Quadrant III: The Respondent Strategies

- Sample survey
 - the investigator tries to obtain evidence that will permit him or her to estimate the distribution of some variables, and/or some relationships among them, within a specified population
- Examples:
 - public opinion surveys on voting intentions, political preferences, buying intentions

Quadrant III: The Respondent Strategies

- The systematic gathering of responses to questions or stimuli formulated by the experimenter
 - in contrast to the observation of behaviors of the participants within an ongoing behavior system

Quadrant IV: The Theoretical Strategies

Quadrant IV: The Theoretical Strategies

- Formal theory
 - the researcher focuses on formulating general relations among a number of variables of interest
- Computer simulation
 - a complete and closed system that models the operation of the concrete system without any real people

Consider when reading

- Does the paper acknowledge the strengths and weaknesses of its evaluation methods?
- Is the research evidence based on only a single evaluation method?

Types of studies

Types of studies

- Three basic forms:
 - Baserates (how often?)
 - Correlations (are properties related?)
 - Differences (comparison or difference)

Base rates

- How often Y occurs in the general case
- Used as a basis for deciding whether the rate of Y in some particular case is “notably” high or low

Correlation

- Is there covariation in the values of two properties or features of system?
- Correlations: High – Low; Positive – Negative; Zero
- Linear or nonlinear relation between variables
- Causation versus correlation!

Difference and comparison

- Is there a difference between X and Y?
- Is there a difference between X and no X?

Validity

Validity

- Five different types of validity
 - Statistical conclusion validity
 - Internal validity
 - Construct validity
 - Measurement validity
 - External validity

Validity

- Statistical conclusion validity:
 - Difference arisen just by chance?
 - Relationship between cause and effect

Validity

- Internal validity:
 - How close can you come to asserting that the presence of X caused the altered level of Y values?
 - Difference in Y associated with a difference in X does not necessarily imply a causal role for X
 - Are there other factors which may covary with X and they, rather than X, might have produced the change in Y

Validity

- Construct validity:
 - How well defined are the theoretical ideas in our study?
 - How clearly understood are the conceptual relations being explored?

Validity

- Measurement validity:
 - Do your measures actually measure the construct of interest?
 - How well have you operationalized the construct?

Validity of Findings

- External validity:
 - How confident you can be that your findings will hold upon replication?
 - How general are your our findings?
 - How confident you can be that you can make predictions about other situations?

Common errors made in research

- Selective observation
- Inaccurate observation
- Overgeneralization
- Made-up information
- Ex post facto hypothesizing
- Illogical reasoning
- Ego involvement in understanding
- Premature closure of inquiry
- Mystification