

# Observations on Empirical Computer Research

---

Walt Heimerdinger

WG10.4 Meeting

Obidos

# Hypotheses are Important

---

- Set the scope of the inquiry
    - Determine what is relevant and what is not
    - Some details always will be omitted
  - Many rationale sections miss the mark
    - Include rationale for effort but not for study decisions
  - “Poor experimental methods impede scientific progress”
-

# Constraints are Just as Important

---

- ❑ Helps understanding of omissions and commissions
  - ❑ Helps potential users decide if results or data are relevant to them
-

# Metrics

---

- Should match the hypothesis
  - Should be measurable
  - Should avoid bias for/against a particular approach
-

# Statistical Studies Require Good Statistics

---

- ❑ Include statistician on the team
  - ❑ Understand what statistical approaches are appropriate
    - Known shortfalls and advantages of alternatives
    - Knowledge of existing libraries
-

# Evaluation Context

---

## □ “Real World”

- Observations of external process
- Constrains type and quality of data
- Often limits ability to distribute data
- Ground truth may not be available

## □ “Artificial World”

- Context is expensive to generate and reproduce
  - Stimuli must be high quality
  - Skill of human participants (e.g. Red Team)
-

# People are VERY Difficult to Include

---

## Demographics

- Match to population of interest
- Availability a major determinant

## Legal / Ethical constraints

- Institutional Research Boards proscribe what can be done to a person
  - Personally Identifiable Information limitations limit distribution of data
-

# Multi-Party Evaluation

---

- Extends evaluation
    - Integrates evaluation target in larger environment
    - Expands evaluation context in possibly unexpected directions
  - Goals must be shared
    - How much independence is productive
    - Can be adversarial or cooperative
  - Expensive
-



# Sharing

---

## Results

- Repeatability may/may not be important
  - Study may be single-point observation

## How are results best communicated

- Short paper advertises results in summary form - assumes follow-up by interested parties
  - Experimental paper includes specific sections on experimental methodology
  - Report with sufficient detail to reproduce the effort
  - Executable framework to recreate and possible extend effort
-

# Artifacts

---

- Data
    - Self-documenting - e.g. with metadata
    - Machine readable
    - Anonymized
  - Methodology
    - Specify controlled and uncontrolled variables
    - Use of accepted standards or libraries can help
  - Framework
    - Flexibility
    - Source code
    - Documentation
      - Users guide
-

# Instrumentation

---

- ❑ Testing and evaluation usually a small part of the budget
    - Data collection usually is expensive in time and effort
  - ❑ Experiment context is part of the data
    - Need to capture metadata
  - ❑ Benchmarks
    - Can provide common context
    - Can save enormous amount of effort
    - Scope of benchmark may not fit
    - Long-term relevance is difficult to maintain
-