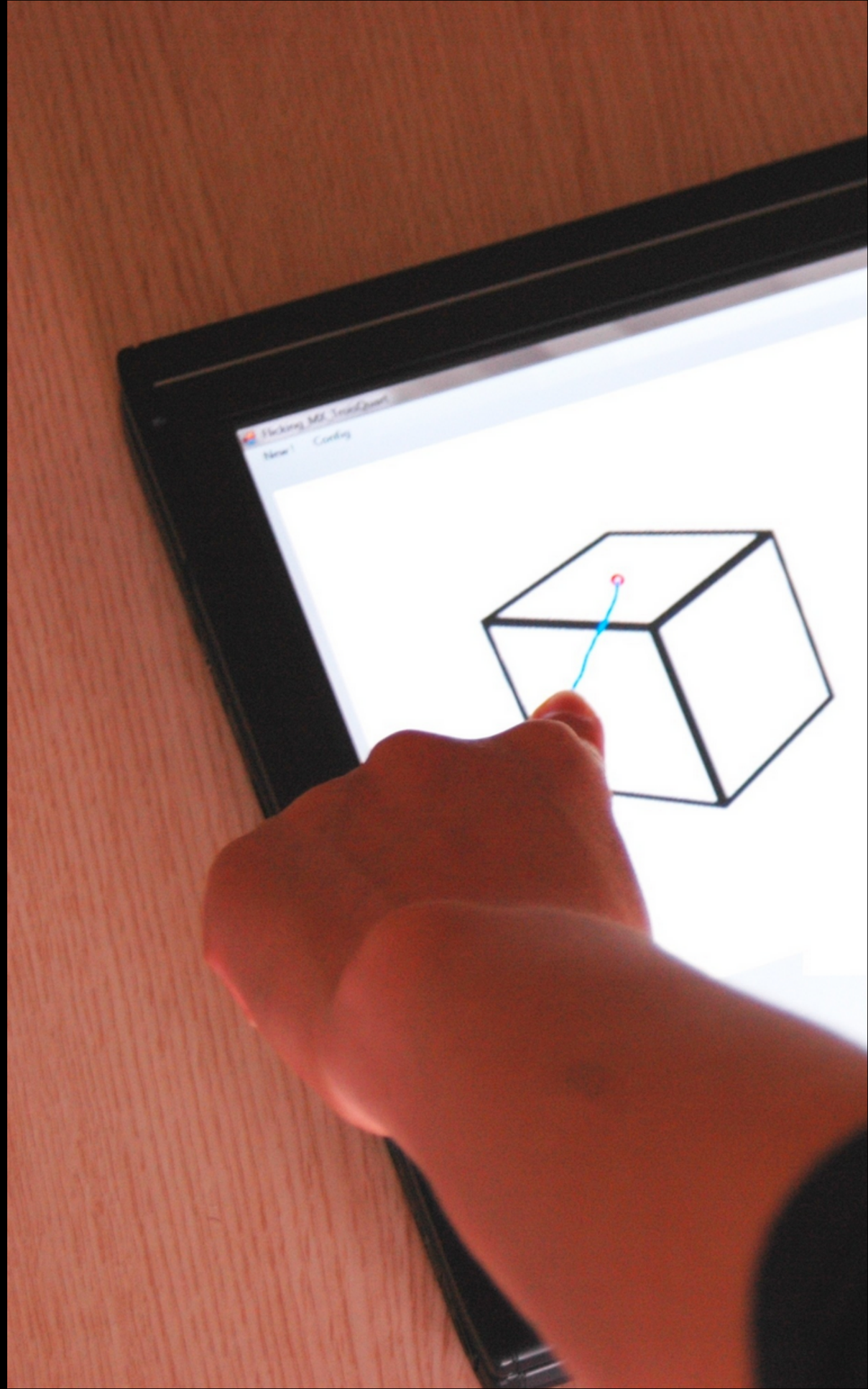
Evaluation & User Study

Byungkyu Kang
FourEyes Lab,
Dept. of Computer Science
UC Santa Barbara



User Study

- What is a study?
  - Empirically testing a hypothesis
- Why run a study?
  - Determine ‘truth’
  - Evaluate if a statement is true
- User Study on Different Platforms
  - Online / Offline

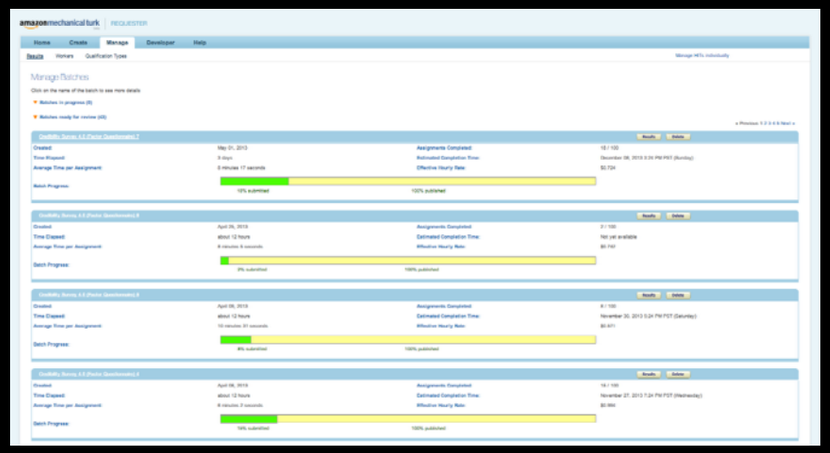
Purpose of User Study

- Evaluate New Interface
- Find the Ground Truth
- Verify a Hypothesis
- Discover errors and areas of improvement

What to Measure?

- Usability Testing (User Study in HCI)
  - **Efficiency** : time and steps in a given task
  - **Accuracy** : mistakes (fatal or recoverable?)
  - **Recall** : How much does the person remember?
  - **Emotional response** : feeling about the task (confident, stressed? recommendable?)

Crowdsourcing User Study



- Online User Study using Crowdsource Platform
  - **General Usability Test**
  - **Ground Truth Annotation**
    - Amazon Mechanical Turk, CrowdFlower
    - Micro-tasks on the Internet
    - Large sample, fast and low cost
- Kittur et al., Crowdsourcing user studies with Mechanical Turk. (CHI '08)

Crowdsourcing User Study

A Type

Interface A

Interface B

Questionnaire

B Type

Screening Task

Task A

Task B

Examples of User Study

- User Interface Design
  - System, Application, Web Search
- User Experience Evaluation
- Virtual or Augmented Reality
- Ground Truth Annotation
- Visualization
  - Scientific Visualization
  - Information Visualization

What does User Study do?

System

Qualitative Performance

User Study

Measure Quantitative Performance

When

Why

How

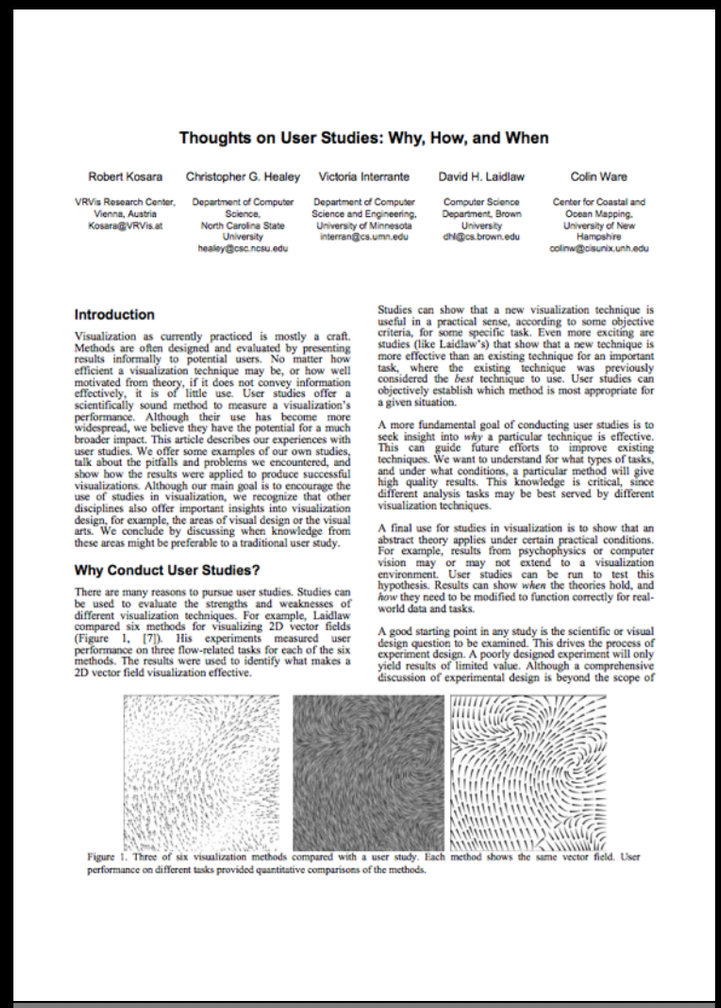
Evaluate Overall Performance

User Study in InfoVis

"User studies offer a scientifically sound method to measure a visualization's performance"

"to evaluate the strengths and weaknesses of different visualization techniques"

Kosara, Robert, et al. "Thoughts on user studies: Why, how, and when." IEEE Computer Graphics and Applications 23.4 (2003): 20-25.



Thoughts on User Studies: Why, How, and When

Robert Kosara — VRVis Research Center, Vienna, Austria, Kosara@VRVis.at

Christopher G. Healey — Department of Computer Science, North Carolina State University, healey@csc.ncsu.edu

Victoria Interrante — Department of Computer Science and Engineering, University of Minnesota, interran@cs.umn.edu

David H. Laidlaw — Computer Science Department, Brown University, dhl@cs.brown.edu

Colin Ware — Center for Coastal and Ocean Mapping, University of New Hampshire, colinw@cisunix.unh.edu

Introduction

Visualization as currently practiced is mostly a craft. Methods are often designed and evaluated by presenting results informally to potential users. No matter how efficient a visualization technique may be, or how well motivated from theory, if it does not convey information effectively, it is of little use. User studies offer a scientifically sound method to measure a visualization's performance. Although their use has become more widespread, we believe they have the potential for a much broader impact. This article describes our experiences with user studies. We offer some examples of our own studies, talk about the pitfalls and problems we encountered, and show how the results were applied to produce successful visualizations. Although our main goal is to encourage the use of studies in visualization, we recognize that other disciplines also offer important insights into visualization design, for example, the areas of visual design or the visual arts. We conclude by discussing when knowledge from these areas might be preferable to a traditional user study.

Why Conduct User Studies?

There are many reasons to pursue user studies. Studies can be used to evaluate the strengths and weaknesses of different visualization techniques. For example, Laidlaw compared six methods for visualizing 2D vector fields (Figure 1, [7]). His experiments measured user performance on three flow-related tasks for each of the six methods. The results were used to identify what makes a 2D vector field visualization effective.

Studies can show that a new visualization technique is useful in a practical sense, according to some objective criteria, for some specific task. Even more exciting are studies (like Laidlaw's) that show that a new technique is more effective than an existing technique for an important task, where the existing technique was previously considered the *best* technique to use. User studies can objectively establish which method is most appropriate for a given situation.

A more fundamental goal of conducting user studies is to seek insight into *why* a particular technique is effective. This can guide future efforts to improve existing techniques. We want to understand for what types of tasks, and under what conditions, a particular method will give high quality results. This knowledge is critical, since different analysis tasks may be best served by different visualization techniques.

A final use for studies in visualization is to show that an abstract theory applies under certain practical conditions. For example, results from psychophysics or computer vision may or may not extend to a visualization environment. User studies can be run to test this hypothesis. Results can show *when* the theories hold, and *how* they need to be modified to function correctly for real-world data and tasks.

A good starting point in any study is the scientific or visual design question to be examined. This drives the process of experiment design. A poorly designed experiment will only yield results of limited value. Although a comprehensive discussion of experimental design is beyond the scope of

Figure 1. Three of six visualization methods compared with a user study. Each method shows the same vector field. User performance on different tasks provided quantitative comparisons of the methods.

IEEE VIS 2013: Panel
Evaluation: How Much Evaluation is Enough?

Panelists:
Min Chen, David Ebert, Brian Fisher, Tamara Munzner

Visual Analytics
&
Cognitive Science

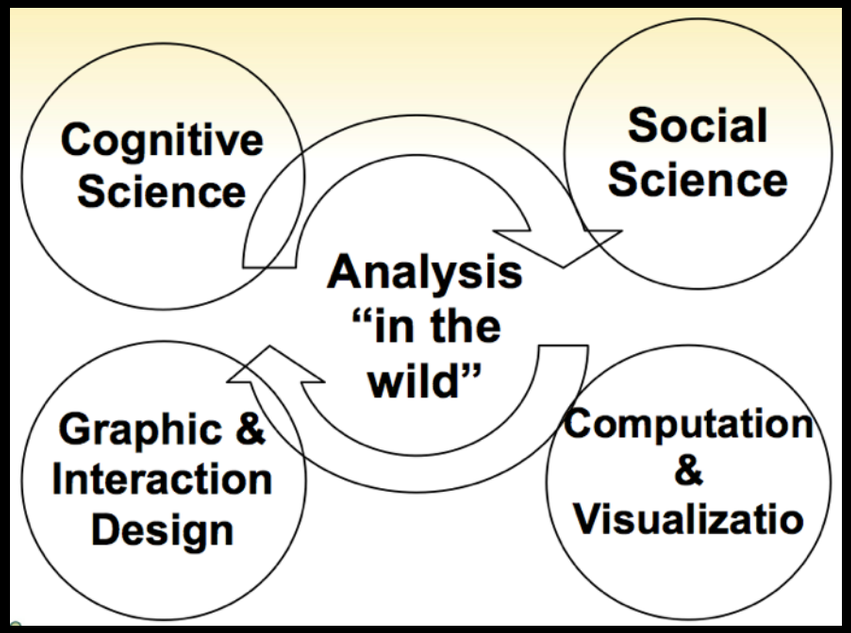
Methodologies

Every paper needs an empirical evaluation?

D-Cog Study

Brian Fischer

- To define an analytics that *underpins analysis*
- Pair analytics : Student "drives", expert "navigates"
  - Student visual analyst & trained domain expert collaborate on analytic task
- Research Snapshot
  - How much of this can we do at VIS?
  - How can we facilitate others to do the rest?
  - How can we interact with them?
  - What organization coordinates the whole process?



Evaluation, When and How

Tamara Munzner

- how to pick the **right evaluation method**.

- A Nested Model for Visualization Design and Validation. Munzner. TVCG 15(6):921-928, 2009 [InfoVis 09]

- Remained Question: do you **need a study** if you're proposing a **new idea**?

Evaluation: broadly interpreted

[A Nested Model for Visualization Design and Validation. Munzner. TVCG 15(6): 921-928, 2009 (Proc. InfoVis 09).]

problem domain:

observe target users using existing tools

data/task abstraction:

encoding/interaction technique:

justify design wrt alternatives

algorithm:

measure system time

analyze computational complexity

analyze results qualitatively

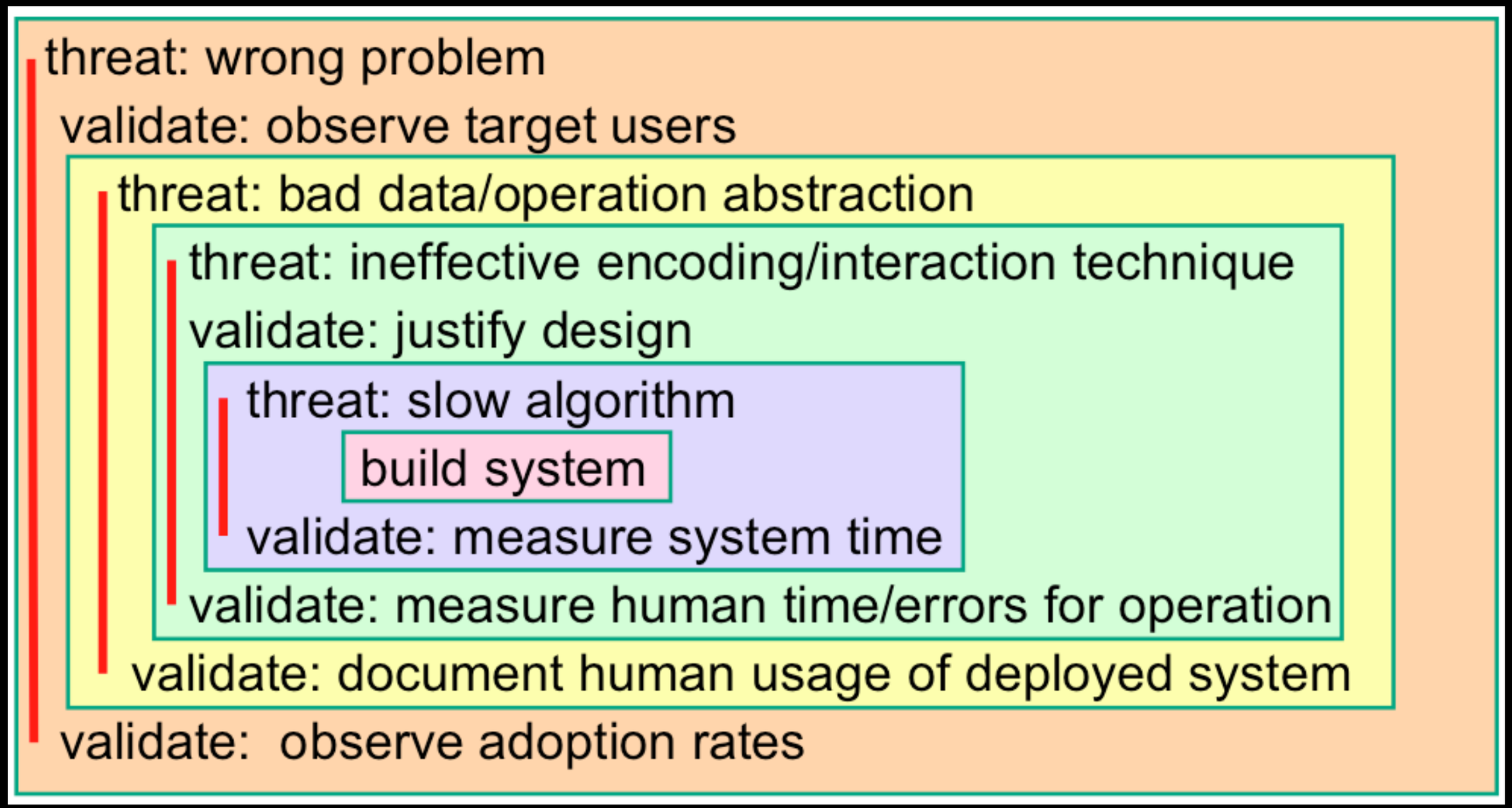
*measure human time with lab experiment ("**user study**")*

observe target users post-deployment ("field study")

measure adoption

http://www.cs.ubc.ca/labs/imager/tr/2009/NestedModel/

Evaluation: broadly interpreted



Threats and validation in the nested model.

Others

David Ebert

- It all depends to context (How much eval?)
  - Answer important questions
    - **Better** than previous contributions?
    - Is the system **effective** and **useful**?
- wrong scientific approach
- statistically significant performance with toy study do not work!
- Publishing without user studies are fine and sometimes better!

Discussions

- Hypothesis “Blinded” vs “Opened”
- Bias-free Design?
- Avoid W.E.I.R.D.(Western, Educated, Industrialized, Rich and Democratic) Society!
- How Many Subjects Required?
- Is Questionnaire Clear or Ambiguous?
- Hawthorne effect (Observer Effect)?