Should Computer Scientists Experiment More?

By Walter Tichy

Outline

- Is Computer Science a Science?
- Why should we experiment?
  - Eight fallacies exposed
- Inherent problems with experimentation

Is Computer Science a Science?

- Tradition: *science deals with fundamental laws of nature*

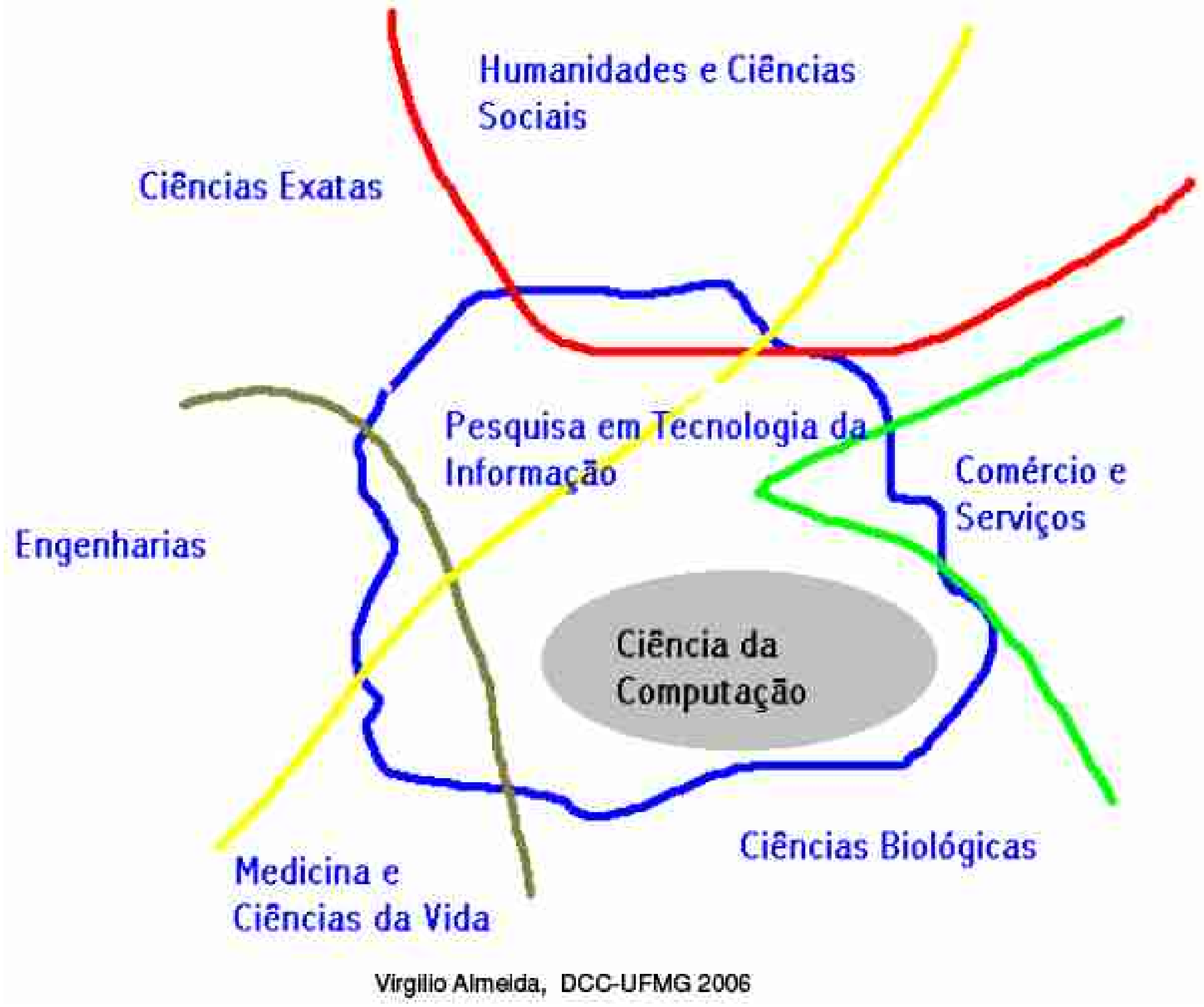
- No, an engineering discipline (Fred Brooks)
  - Computers and programs are human creations
  - Computer science is *not* a natural science

- Yes, much more than synthetic results
  - Study of information structures & processes
  - Synthetic results (computers & programs) are models
  - Difference from traditional science: work with information – neither energy nor matter

Is Computer Science a Science? (P. Denning, CACM 4/2005)

- Science or technology (man-made objects)?
  - Scientific paradigm (Francis Bacon): process of forming hypotheses and testing them through experiments
  - Science means explaining, modeling and predicting phenomena in the world.
  - Computer Science: science of information processes and their interactions with the world.
- Computer Science:
  - Art, science, fundamental principles that are non-obvious and has a future (relations with other areas)
  - But: credibility problem
    - In a sample of 400 papers before 1995, Tichy found that about 50% of those proposing models or hypothesis did not test them. In other fields of science the fraction of papers with untested hypothesis was about 10%.

Futuro: Transdisciplinariedade



Why should we experiment?

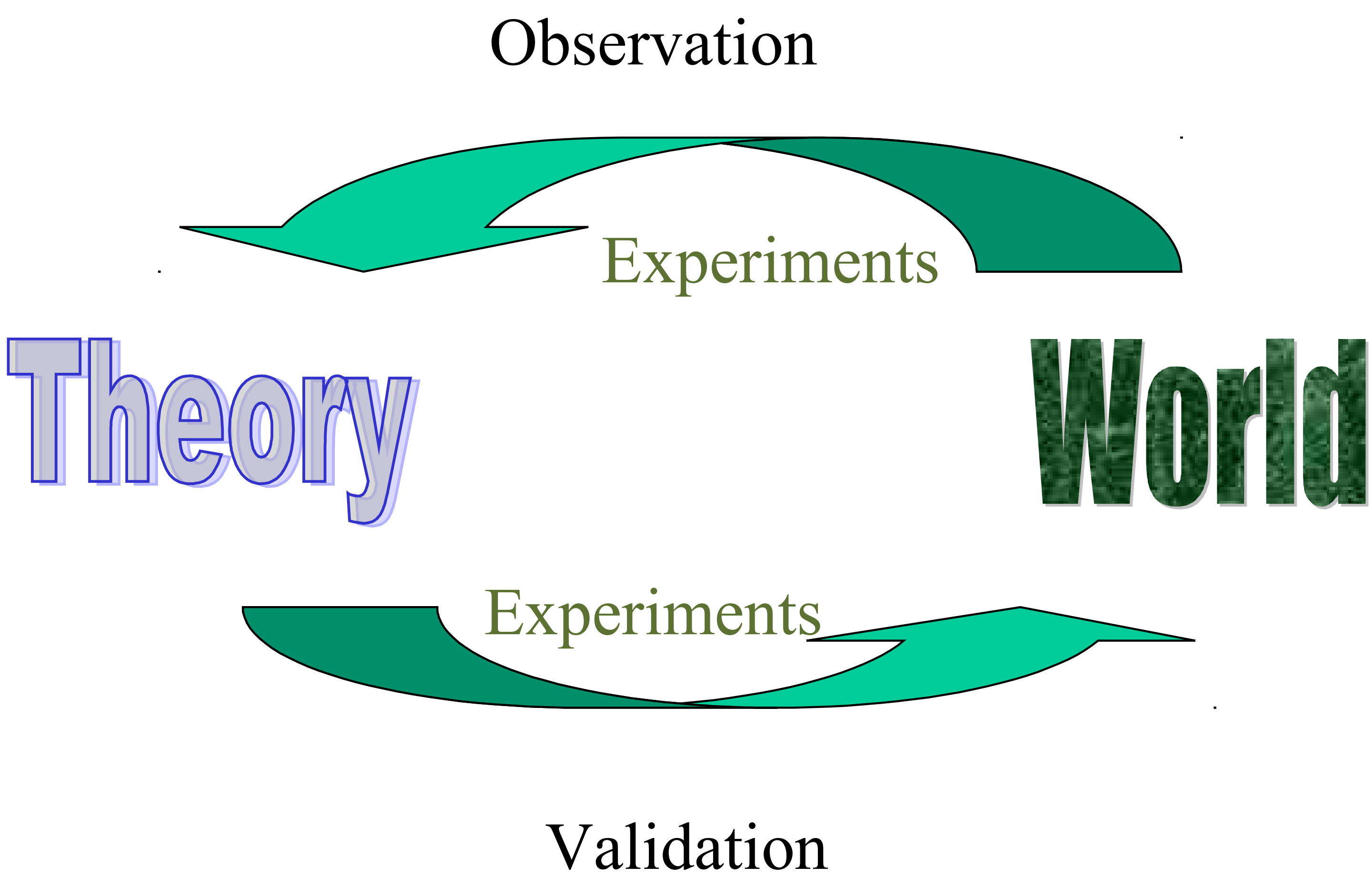
- Theory testing and exploration
  - Theory falsification: falsifiability (or refutability or testability) is the logical possibility that an assertion can be shown false by an observation or a physical experiment.
- Aid with induction or theory derivation
  - Exploration, deriving theories from observation

Eight Fallacies Exposed (#1)

Traditional scientific method isn't applicable

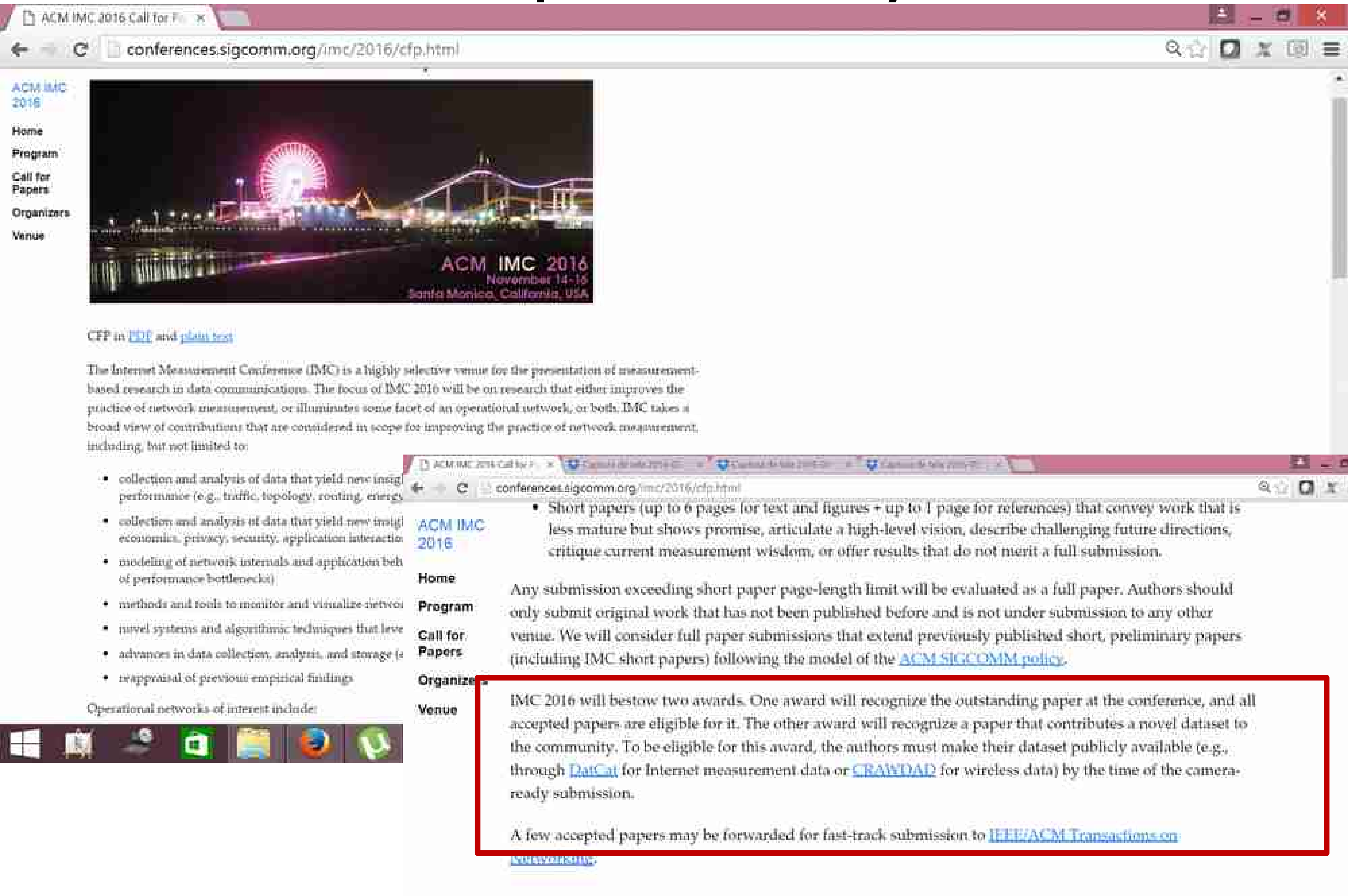
- *Rebuttal: To understand info processes, computer scientists must observe phenomena, formulate explanations, and test them. This <u>is</u> the scientific method.*
- Many CS theories have not been tested
  - OO programming improves programmer productivity, program quality.
- Repeatability is key requirement to any experiment

Eight Fallacies Exposed (#1)



Apresentacao derivada dos slides originais de Virgilio Almeida

Repeatability



Repeatability

Dear Jussara Almeida,

If you are receiving this email is because you don't have a conflict with the two finalists for the best dataset award of PAM 2009. I've asked all candidates to send me the link to their datasets; many responded that they can't publish their data, so I've removed them from the list. From the remaining five papers, I've selected the following two papers based on your previous votes and the reviews. Now, we need to pick the winner! :)

Please send me your vote by March 18th.

Repeatability

… I recommend the authors to make their code publicly available

(comment by reviewer of Elsevier Computer Communications)

Eight Fallacies Exposed (#2)

The current level of experimentation is good enough

- *Rebuttal: Relative to other sciences, the data shows that computer scientists validate a smaller percentage of their claims*
- 40-50% of software papers published in ACM journals were unvalidated (~ 15% in other areas)
  - Computer scientists publish a lot of untested ideas or that the ideas published are not worth testing
- Balancing theory and engineering with experiment
  - Build reliable base & reduce uncertainties
  - Lead to new areas of investigation
  - Accelerate progress by pruning fruitless approaches

Eight Fallacies Exposed (#3)

Experiments cost too much

- *Rebuttal: Meaningful experiments can fit into small budgets; expensive experiments can be worth more than their cost*
- Constrained by cost
  - Probe importance of research question
  - Plan appropriate research programs
  - Look for *affordable experimental techniques*
  - Intermediate steps with *partial results*
- Industry is beginning to value experiments
  - Three to five year lead over competition
- Experiments in other areas
  - Pharmaceuticals, aeronautics, biology

Eight Fallacies Exposed (#4)

Demonstrations will suffice

- *Rebuttal: Demos can provide incentives to study a question further. Too often, however, these demos merely illustrate a potential*
- Proof of concept
- No solid evidence
- Solid evidence requires careful analysis involving experiments, data and replication
  - Experiments require clear question, experimental apparatus to test the question, data collection, interpretation, sharing of results

Eight Fallacies Exposed (#5)

There's too much noise in the way

- *Rebuttal: Fortunately, benchmarking can be used to simplify variables and answer questions*
- Benchmarks
  - Task domain sample executed by a computer or by a human and a computer
  - allow repeatable and objective comparisons
  - aids in identifying promising approaches and discarding poor ones
  - Benchmark composition is the big challenge
- Experiments involving humans also repeatable (medicine & psychology)

Eight Fallacies Exposed (#6)

Experimentation will slow progress

- *Rebuttal: Increasing the ratio of papers with meaningful validation has a good chance of actually accelerating progress*
- Good conceptual papers will continue to be published
- Need to get beyond assertion
  - “it seems intuitively obvious”,
  - “it looks like a good idea”,
  - “I tried it on a small example and it worked”

Eight Fallacies Exposed (#7)

Technology changes too fast

- *Rebuttal: If a question becomes irrelevant quickly, it is too narrowly defined and not worth spending a lot of effort on.*
- Probe for fundamental and not the ephemeral
- Scientists should anticipate changes in assumptions and proactively employ experiments to explore consequence of changes

Eight Fallacies Exposed (#8)

You'll never get it published

- *Rebuttal: Smaller steps are still worth publishing because they improve our understanding and raise new questions*
- Non-theoretical journals and conferences accept papers on solid experimentation
- Respectable experimentalists articulate how their systems contribute to our knowledge
  - "Systems come and go. We need insights about the concepts and phenomena underlying such systems"

41st International Conference on

VERY LARGE DATA BASES

Hilton Waikoloa Hotel • Kohala Coast, Hawai'i
August 31 - September 4, 2015

Search

General Information
- Conference Overview
- Conference Officers
- PVLDB Review Board
- Acknowledgments

Program
- Schedule at a Glance
- Full Program
- Talk by PC Chairs
- Accepted Research Papers
- Accepted Industrial Papers
- Turing Award Lecture
- Accepted Demos
- Accepted Tutorials
- Keynotes
- Panels
- Workshops
- PhD Workshop
- Social Events
- **Poster Sessions**
  - Poster Session I
  - Poster Session II
- Session Guidelines
- Awards
- Booklet

Experiments & Analysis Papers

Green turtle, Chelonia mydas, Hawaii, Photo by Brocken Inaglory, GNU Free Documentation/Creative Commons Attribution license, via Wikimedia

Experiments and Analysis Papers focus on the experimental evaluation of existing algorithms, data structures, and systems. Papers proposing new techniques should continue to be submitted to the regular research track. The primary contribution of Experiments and Analysis papers is performance evaluation through analytical modeling, simulation, and/or experiments. Suitable papers can fit in different categories:

1. **Experimental Surveys**: papers that compare a wide spectrum of approaches to a problem and, through extensive experiments, provide a comprehensive perspective on the results available and how they compare to each other.
2. **Result Verification**: papers that verify or refute results published in the past and that, through a renewed performance evaluation, help to advance the state of the art.

2015 VLDB Swag

Follow me on Twitter

Tweets by @VLDB2015

VLDB 2015

The thirteenth issue of the eighth volume of the

Inherent Problems with Experimentation

- Unrealistic assumptions
- Manipulated data
- Impossible to quantify key variable
- Few *Competing Theories* in CS
  - Rarely produce falsifiable theories: tend to pursue math theories that are disconnected from the real world

Similar assessment, 10 years later

- Comm. of ACM Nov 2007
  - “In computer science, theory seems to play a more dominant role, typically with little if any direct connection to experimentation”
  - Experimentation:
    - Observe, measure under controlled conditions, replicate previous work
    - Systems are increasingly complex and require objective measurement to be studied and understood
    - “New system designs must be evaluated in the context of representative workloads”
    - “Evaluations under controlled conditions are critical for progress in the development of practical systems”

Experimental Computer Science: The Need for A Cultural Change

- Science
  - Observation + Hypothesis Testing + Reproducibility

- Research in CS:
  - typically part of
    - Engineering (building tools/systems) or
    - Math (building and studying abstract processes/structures)
  - Counter-example: Self-similarity in computer network traffic
    - Originated from pure observation - SCIENCE

The role of experimentation

Experimentation in science Experimental feedback in engineering

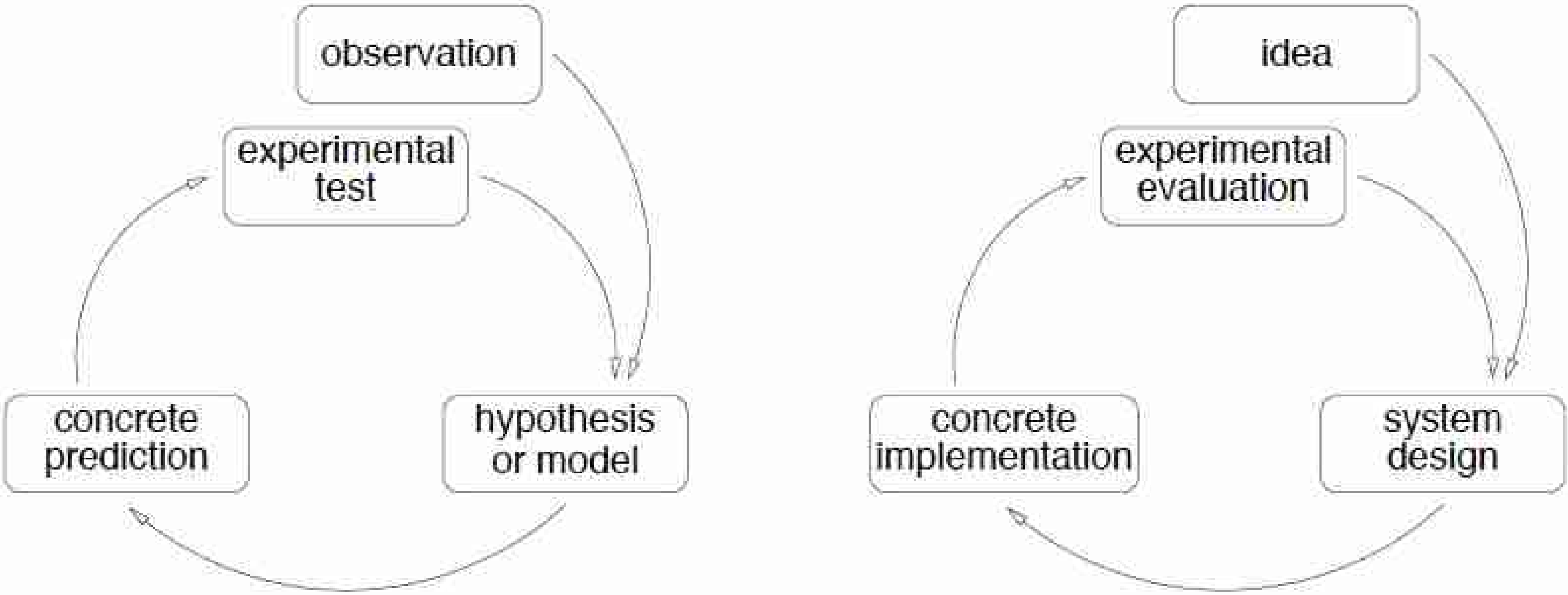


Figure 1: *A comparison of the scientific method (on the left) with the role of experimentation in system design (right).*

"In many cases system experiments are more demonstrations that the idea or system works than a real experiment

The role of experimentation

- Key questions that should be made:
  - Is it based on empirical evaluation and data?
  - Was the experiment designed correctly?
  - Is it based on a toy or a real situation?
  - Were the measurements used appropriate for the goals of the experiment?
  - Was the experiment run for a long enough time?

- Moreover, need to avoid (when comparing against alternative):
  - Bias in favor of own system
  - Tendency to compare against restricted, less optimized versions of the competition

Observation+Hypothesis+ Reproducibility

- Observation (measurement) leads to models and better understading of process at hand

  "Art and science have their meeting point in method"
- Need appropriate **metrics**
- Models turn measured data into **information**
  - Good models are simple (do we need a new model?)
  - Do not fit the measurements to the theory but rather fit the theory to the measurements
- Actual numbers may be of little use, the **understanding** derived from them has wider applicability