Viewpoints to Experimental CS

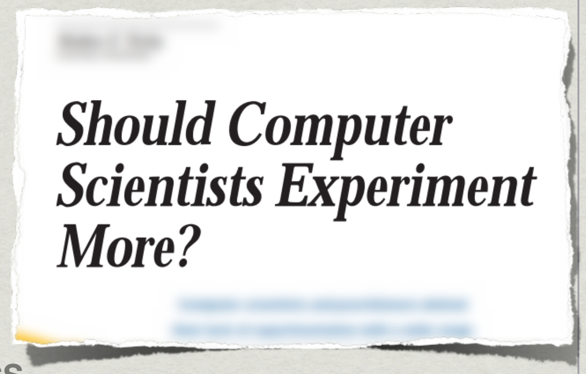
Matti Tedre, Stockholm University
matti.nospam.tedre@acm.nospam.org

A Very Short History

- Feldman & Sutherland (1979) *Rejuvenating Experimental Computer Science*
- *ACM Executive Committee position on the crisis in experimental computer science* (1979)
- *Starting from the 1980s, hundreds of arguments on experimental computer science*
- Tichy (1998) *Should Computer Scientists Experiment More?*

Outline

1. More
2. Computer scientists
3. Experiment
4. Should
5. Challenges for Experimental CS


Should Computer Scientists Experiment More?

1. Should Computer Scientists Experiment More?

How much do we actually experiment?

How Much Do We Experiment? (1/2)

- According to one (old) study:
  - 40% of articles about new designs and models **lack** experimentation
  - 50% of SE articles

TICHY ET AL., CACM 23(10):544

How Much Do We Experiment? (2/2)

	Research Method	CS	SE	IS
AR	Action Research	-	0%	0.8%
FE	Field Experiment	-	<1%	1.6%
FS	Field Study	0.2%	<1%	24.5%
GT	Grounded Theory	-	<1%	0.2%
HE	Hermeneutics	-	<1%	-
ID	Instrument Development	-	-	3.5%
LH	Laboratory Experiment - Human Subjects	1.8%	3.0%	16.2%
LR	Literature Review/analysis	.3%	1.1%	0.8%
LS	Laboratory Experiment - Software	1.9%	<1%	0.6%

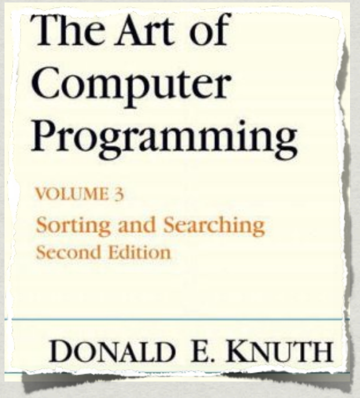
GLASS ET AL., CACM 47(6):91

Where Should We Seek for Experiments?

> The hypothesis may concern a law of nature—for example, one can test whether a hashing algorithm's average search time is a small constant independent of the table size by

DENNING, CACM 23(10):544

Pioneers in Sorting + Searching



- Where's the experimentation?
- Are we really justified to say "all work in computing needs experimentation"?
- Need to look closer to what people mean by "computer science"

KNUTH, TAOCP, VOL.3

2. Should Computer Scientists Experiment More?

What do we mean by “computer science” here?

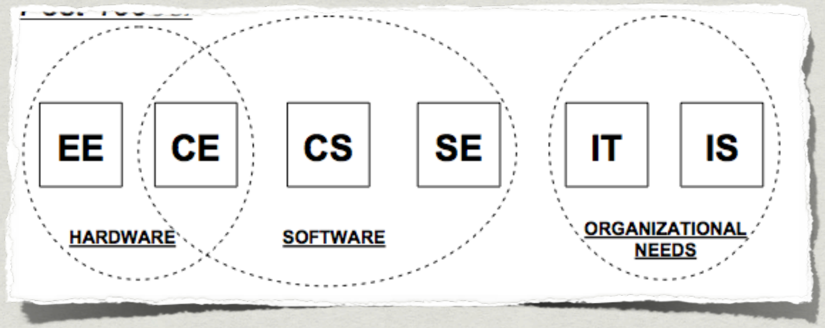
Is Computing a Science?

No!	Dijkstra (1987), McKee (1995), Brooks (1996), Hoare (early works)
Yes.	Denning (2007), Simon (1969), Newell et al. (1969), Rosenbloom (2004), Ralston & Shaw (1980), Minsky (1979), McCracken (1979), ...
Yes, but...	Knuth (2001), Tichy (1998), Hartmanis (1993/1994), Vessey et al. (2002), Fletcher (1995)

...If So, What Is It a Science of?

Computers	Hamming (1969)
Computers + related phenomena	Newell et al. (1969)
Algorithms + related phenomena	Knuth (1974)
Information processing	Forsythe (1967)
Complexity	Simon (1969)
Classes of computations	Dijkstra (1972)
Programming	Khalil & Levy (1978)
Information processes & transf.	Denning et al. (1981)

How About the Computing Fields?



ACM/IEEE CURRICULA (2005)

How Do Computer Scientists Work?

- Mathematical analysis
- Conce
- Conce
- Surve
- Scree
- Effect
- Bench
- Simula
- Emula
- Participatory design
- Hermeneutics
- Dynamic analysis
- Simulation
- Legacy data analysis
- Lessons learned
- Static analysis
- Project monitoring
- Focus groups
- Content analysis
- Ethnography
- Grounded theory
- Critical theory
- Incremental development
- Performance analysis

What is the role of experiments in computer science methodology?

3. Should Computer Scientists **Experiment** More?

What do computer scientists mean by "experiment"?

5 Views to Experiments in Computer Science

1. Demonstration experiment
2. Trial experiment
3. Field experiment
4. Comparison experiment
5. Controlled experiment

"EXPERIMENT"?!?!

1. Demonstration Experiment

- It's not known if task *t* can be automated efficiently / reliably / feasibly / cost-efficiently etc.
- A demonstration of *experimental* technology shows that it can be done
  - ..."experiment"?

2. Trial Experiment

- It's not known how well system *p* meets its specifications / performs
- A trial evaluation is designed to *experiment* ("test") with the qualities of the product
- Can be lab-based or in the use environment
  - E.g., performance analysis (w/o comparison)
  - Benchmarking (Gustedt et al., 2009)

3. Field Experiment

- It's not known how well a system works in its sociotechnical context
- Common in information systems (Palvia et al., 2003)
- Offers more control than case studies and surveys; typically a quasi-experiment or limited-control experiment

4. Comparison Experiment

- It is not known if algorithm A outperforms B with data set *d* and parameters *p*
- An experiment is set up to measure and compare A(*d*,*p*) and B(*d*,*p*)
- Typical of incremental development work where the aim is to do task *x* better
- Typically doesn't follow the blinding principle

5. Controlled Experiment

- It is not known if *x* and *y* are associated, or if *x* causes *y*.
- The gold standard of scientific work
- Enables generalization and prediction

How do we justify saying “We SHOULD experiment more”?

4. Should Computer Scientists Experiment More?

Argumentum ad Antiquitatem

- One of the most famous arguments for increasing experimentation in computing (Tichy et al., 1995)

10 J. SYSTEMS SOFTWARE
1995; 28:9–18

For comparison, we reviewed publications from two other fields: volume 5 (1993) of *Neural Computation* (*NC*), and numbers 1 and 3 of volume 33 (1994) of *Optical Engineering* (*OE*). *NC*. published

TICHY ET AL., J. SYST. SOFTW. 28:9–18

1) How ysics?

Computing journals

Am. J. of Physics

Method	%	J2
Not applicable		2
No experimentation	29.7	58%
Replicated	1.1	5%
Synthetic	2.1	
Dynamic analysis	1.2	5%
Simulation	5.5	
Project Monitoring	0.2	
Case study	10.3	16%
Assertion	34.2	4%
Field study	1.2	
Literature search	3.0	11%
Legacy data	1.9	
Lessons learned	8.7	5%
Static analysis	0.7	
Yearly totals		21

- Conclusion: we should experiment *less*?

ZELKOWITZ & WALLACE, INF. & SW TECH. 39:735–743

2) How Unique is Computer Science?

- Maybe computing is just different?
- Compare*...
  - The work that leads to Nobel Prize in Physics
  - The work that leads to Turing Awards
- If it’s a unique field, does it not have unique ways of working, too?

*HARTMANIS, PROC. 13TH CONF. OFOST: 1–12

5. Challenges for Experimental CS

Challenge 1: Sloppy terminology

- experimentation terminology is sloppy.

* But computing has always been liberal with terms!

ZELKOWITZ & WALLACE, IEEE COMPUTER 31(5):23–31

Challenge 2: Empirical vs. Experimental

* When we say "experimental", do we really mean "empirical"?

Challenge 3: Develop Experimental Protocols

1. No precautions are taken against experimenter bias. Most of the experimental hypotheses are of

* In many branches of CS there **are** standards for data, parameters, and measurement!

FLETCHER, J. SYS. SOFTW. 30(1995):161–163

Challenge 4: Tacked-on "Hypotheses"

creativity of programmers. The basic hypothesis of the ARPAnet is that long distance message and file transfer services would significantly increase scientific productivity by permitting critical masses of researchers to form across long distances. These

- CS projects can be great even if they aren't based on hypothesis testing!

DENNING, CACM 23(10):544

Challenge 5: "Computer Science"

- When we say "CS" do we really mean "SE", or "Health informatics" or "educational technology"?

Challenge 6: Arguments to Other Disciplines

- Is it really necessary to compare with physics?
  - Why not astronomy?
  - ...mathematics?
  - ...economics?

Challenge 7: Language from the 1600s

- In our arguments, why are we still referring to ideas about science developed in the 1600s?
- We nowadays know much better how science works.

Thanks!

Questions, comments?
Obvious errors?
matti.nospam.tedre@acm.nospam.org