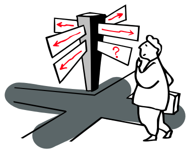
Research Methods in Computer Science

(Serge Demeyer — University of Antwerp)

Lab on Reengineering
http://lore.ua.ac.be/

Universiteit Antwerpen

Research Methods



Introduction
- Origins of Computer Science
- Research Philosophy

Research Methods
- Feasibility study
- Pilot Case
- Comparative study
- Literature survey
- Formal Model
- Simulation

Conclusion
- Studying a Case
  vs. Performing a Case Study
  + Proposition
  + Unit of Analysis
  + Threats to Validity

Computer Science

All science is either physics or stamp collecting (E. Rutherford)

We study artifacts produced by *humans*

Computer science is no more about computers than astronomy is about telescopes. (E. Dijkstra)

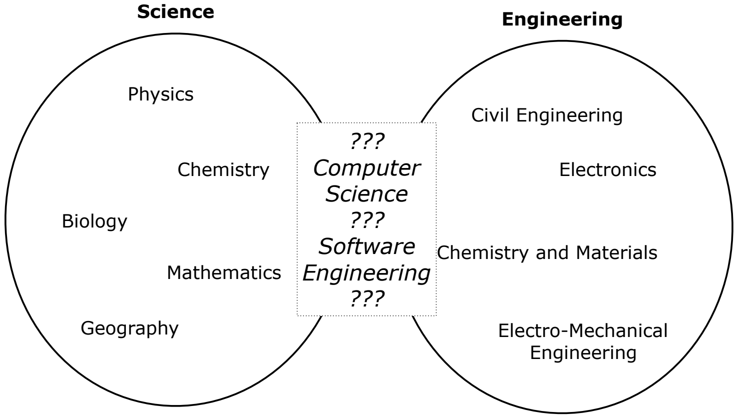
Computer science

Computer engineering

Informatics

Software Engineering

Science vs. Engineering



Influence of Society

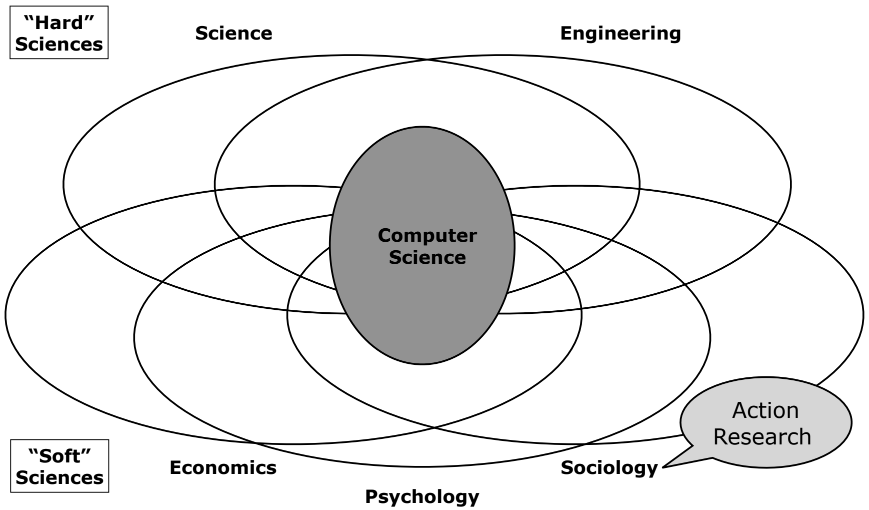
Lives are at stake
(e.g., automatic pilot, nuclear power plants)

Huge amounts of money are at stake
(e.g., Ariane V crash, Denver Airport Baggage)

Software became Ubiquitous
… it's not a hobby anymore

Corporate success or failure is at stake (e.g., telephone billing, VTM launching 2nd channel)

Interdisciplinary Nature



The Oak Forest
Robert Zünd - 1882

Franz and Luciano
Franz Gertsch - 1973

Dominant view on Research Methods

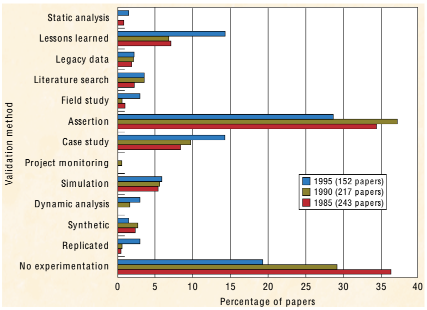
Physics
("The" Scientific method)
- form hypothesis about a phenomenon
- design experiment
- collect data
- compare data to hypothesis
- accept or reject hypothesis
  - + … publish (in Nature)
- get someone else to repeat experiment (replication)

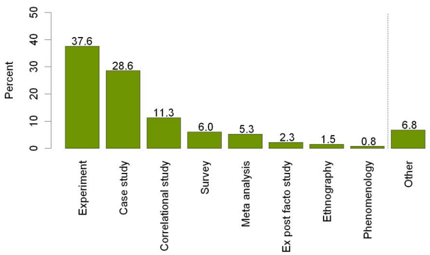
Medicine
(Double-blind treatment)
- form hypothesis about a treatment
- select experimental and control groups that are comparable except for the treatment
- collect data
- commit statistics on the data
- treatment ⇒ difference (statistically significant)

Cannot answer the "big" questions
… in timely fashion
- smoking is unhealthy
- climate change
- darwin theory vs. intelligent design
- …
- agile methods

Research Methods in Computer Science

Different Sources
- Markin V. Zelkowitz and Dolores R. Wallace, "Experimental Models for Validating Technology", IEEE Computer, May 1998.
- Easterbrook, S. M., Singer, J., Storey, M, and Damian, D. Selecting Empirical Methods for Software Engineering Research. Appears in F. Shull and J. Singer (eds) "Guide to Advanced Empirical Software Engineering", Springer, 2007.
- Gordona Dodif-Crnkovic, "Scientific Methods in Computer Science"
- Andreas Höfer, Walter F. Tichy, Status of Empirical Research in Software Engineering, Empirical Software Engineering Issues, p. 10-19, Springer, 2007.





Case studies - Spectrum

"case studies" are widely used in computer science
⇒ "studying a case" vs. "doing a case study"

Simulation
- what if ?

Formal Model
- underlying concepts ?

Literature survey
- what is known/unknown ?

Comparative study
- is it better ?

Pilot Case, Demonstrator
- is it appropriate ?

Feasibility study
- is it possible ?

Source: Personal experience
(Guidelines for Master Thesis Research – University of Antwerp)

The sixteenth of september
Rene Margritte

Feasibility Study

Here is a new idea, is it possible ?

➡ Metaphor: Christopher Columbus and shorter route to India

- Is it *possible* to solve a specific kind of problem … effectively ?
  + computer science perspective (P = NP, Turing test, …)
  + engineering perspective (build efficiently; fast — small)
  + economic perspective (cost effective; profitable)
- Is the technique new / novel / innovative ?
  + compare against alternatives
    ➡ See literature survey; comparative study

- Proof by construction
  + build a prototype
  + often by applying on a "case"

- Conclusions
  + primarily qualitative; "lessons learned"
  + quantitative
    - economic perspective: cost - benefit
    - engineering perspective: speed - memory footprint

The Prophet
Pablo Gargallo

Pilot Case (a.k.a. Demonstrator)

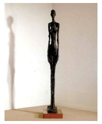
Here is an idea that has proven valuable; does it work for us ?

- proven valuable
  + accepted merits (e.g. "lessons learned" from feasibility study)
  + there is some (implicit) theory explaining why the idea has merit
- does it work for us
  + context is very important

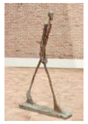
- Demonstrated on a simple yet representative "case"
  + "Pilot case" ≠ "Pilot Study"

- Proof by construction
  + build a prototype
  + apply on a "case"

- Conclusions
  + primarily qualitative; "lessons learned"
  + quantitative; preferably with predefined criteria
    ➡ compare to context before applying the idea !!



Walking man
Standing Figure
– Alberto Giacometti



Comparative Study

Here are two techniques, which one is better for a given purpose ?

- (Not necessarily absolute ranking)
- Where are the differences ? What are the tradeoffs ?

- Criteria check-list
  - \+ predefined
    - \- should not favor one technique
  - \+ qualitative and quantitative
    - \- qualitative: how to remain unbiased ?
    - \- quantitative: represent what you want to know ?
  - \+ Criteria check-list should be complete and reusable !
    - ➡ See literature survey

- Score criteria check-list
  - \+ Often by applying the technique on a "case"

- Conclusions
  - \+ compare: typically in the form of a table

Torben Giehler
Matterhorn

Paul Klee
Niesen

Literature Survey

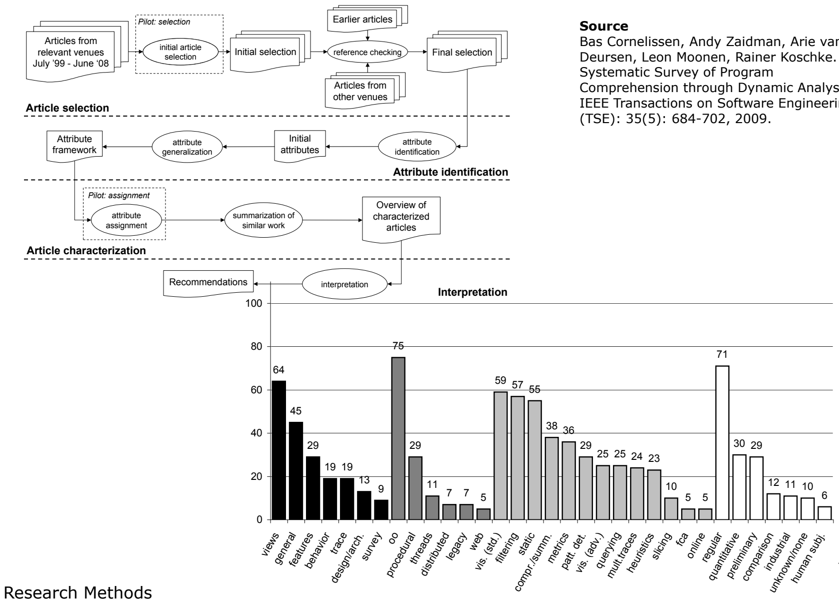
What is known ? What questions are still open ?

- source: B. A. Kitchenham, "Procedures for Performing Systematic Reviews", Keele University Technical Report EBSE-2007-01, 2007

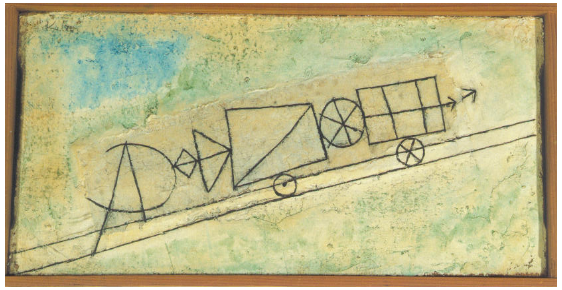
Systematic

- "comprehensive"
    - ➡ precise research question is prerequisite
  - \+ defined search strategy (rigor, completeness, replication)
  - \+ clearly defined scope
    - \- criteria for inclusion and exclusion
  - \+ specify information to be obtained
    - \- the "cases" are the selected papers
- outcome is organized
  - \+ classification
    - ➡ table
  - \+ taxonomy
    - ➡ tree
  - \+ conceptual model
    - ➡ frequency

Literature survey - example



Source
Bas Cornelissen, Andy Zaidman, Arie van Deursen, Leon Moonen, Rainer Koschke. A Systematic Survey of Program Comprehension through Dynamic Analysis IEEE Transactions on Software Engineering (TSE): 35(5): 684-702, 2009.



Klee
Bergbahn

Vojin Bakic
Bull

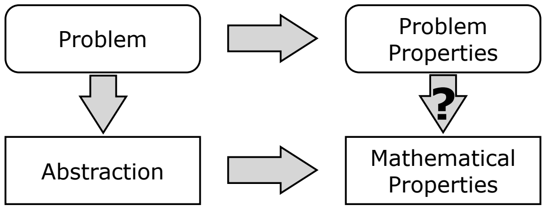
Formal Model

How can we understand/explain the world ?

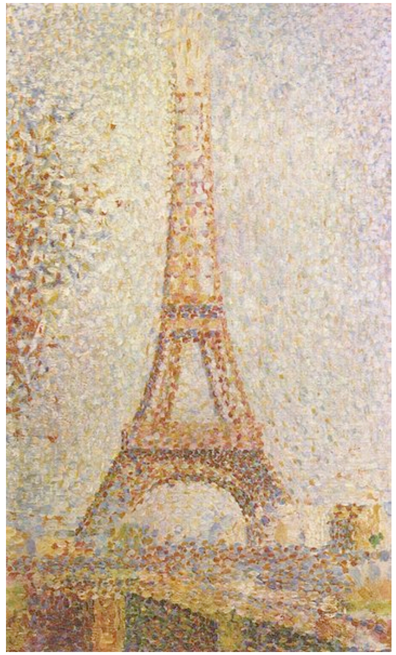
- make a mathematical abstraction of a certain problem
  - + analytical model, stochastic model, logical model, re-write system, ...
  - + often explained using a “case” (toy example)
- prove some important characteristics
  - + based on inductive reasoning, axioms & lemma’s, …

Motivate

- which factors are irrelevant (excluded) and which are not (included) ?
- which properties are worthwhile (proven) ?
  - ➡ See literature survey



Hodler
Eiger, Mönch and Jungfrau in the Morning Sun



Seurat
Eiffel Tower

Simulation

What would happen if … ?

- study circumstances of phenomena in detail
  + simulated because real world too expensive; too slow or impossible
- make prognoses about what can happen in certain situations
  + test using real observations, typically obtained via a "case"

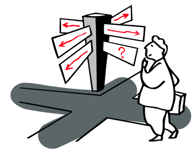
Motivate

- which circumstances are irrelevant (excluded) and which are not (included) ?
- which properties are worthwhile (to be observed/predicted) ?
  ➡ See literature survey

Examples

- distributed systems (grid); network protocols
  + too expensive or too slow to test in real life
- embedded systems — simulating hardware platforms
  + impossible to observe real clock-speed / memory footrpint / …
  ➡ Heisenberg uncertainty principle

What did we do ?



Introduction
- Origins of Computer Science
- Research Philosophy

Research Methods
- Feasibility study
- Pilot Case
- Comparative study
- Literature survey
- Formal Model
- Simulation

Conclusion
- Studying a Case
  vs. Performing a Case Study
  + Proposition
  + Unit of Analysis
  + Threats to Validity

Sources
- Robert K. Yin. Case Study Research: Design and Methods. 3rd Edition. SAGE Publications. California, 2009.
- Bent Flyvbjerg, "Five Misunderstandings About Case Study Research." Qualitative Inquiry, vol. 12, no. 2, April 2006, pp. 219-245.
- Runeson, P. and Höst, M. 2009. Guidelines for conducting and reporting case study research in software engineering. Empirical Softw. Eng. 14, 2 (Apr. 2009), 131-164.

Generalization

Misunderstanding 2:

- One cannot generalize on the basis of an individual case; therefore the case study cannot contribute to scientific development.
  [Bent Flyvbjerg, "Five Misunderstandings About Case Study Research."]

- Understanding
  + The power of examples
  + Formal generalization is overvalued
    - dominant research views of physics and medicine
- Counterexamples
  + one black swan falsifies "all swans are white"
    - case studies generate deep understanding; what appears to be white often turns out to be black
- sampling logic vs. replication logic
  + sampling logic: operational enumeration of entire universe
    - use statistics: generalize from "randomly selected" observations
  + replication logic: careful selection of boundary values
    - use logic reasoning: presence of absence of property has effect
    ➡ Requires precise propositions

Spectrum of cases

Feasibility study

created for *explanation*
- foo, bar examples
- simple model; illustrates *differences*

Toy-example

accepted teaching vehicle
- "textbook example"
- simple but illustrates *relevant* issues

Exemplar

Martin S. Feather , Stephen Fickas , Anthony Finkelstein , Axel Van Lamsweerde, Requirements and Specification Exemplars, Automated Software Engineering, v.4 n.4, p. 419-438, October 1997

Case study

Runeson, P. and Höst, M. 2009. Guidelines for conducting and reporting case study research in software engineering. Empirical Softw. Eng. 14, 2 (Apr. 2009), 131-164.

real-life example
- industrial system, open-source system
- context is difficult to grasp

Case

Mining Software Repositories Challenge.
[Yearly workshop where research tools compete against one another on a common predefined case.]

competition (tool oriented)
- approved by community
- *comparing*

Community case

Susan Elliott Sim, Steve Easterbrook, and Richard C. Holt. Using Benchmarking to Advance Research: A Challenge to Software Engineering, Proceedings of the Twenty-fifth International Conference on Software Engineering, Portland, Oregon, pp. 74-83, 3-10 May, 2003.

Benchmark

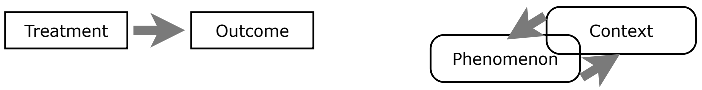
benchmark
- approved by community
- known context
- "*planted*" issues

Case study — definition

Definition

- *A case study is an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between the phenomenon and context are not clearly evident*
  [Robert K. Yin. Case Study Research: Design and Methods; p. 13]

  + empirical inquiry: yes, it is empirical research
  + contemporary: (close to) real-time observations
    - incl. interviews
  + boundaries between the phenomenon and context not clear
    - as opposed to "experiment"



➡ many more variables than data points
➡ multiple sources of evidence; triangulation
➡ theoretical propositions guide data collection
(try to confirm or refute propositions with well-selected cases)

Case study - steps

1. Questions
- most likely "How" and "Why"; also sometimes "What"

2. Propositions (a.k.a. Purpose)
- explanatory: where to look for evidence
- exploratory: rationale and direction
  + example: Christopher Columbus asks for sponsorship
    - Why three ships (not one, not five) ?
    - Why going westward (not south ?)
- role of "Theories"
  + possible explanations (how, why) for certain phenomena
    ➡ Obtained through literature survey

3. Unit(s) of analysis
- What is the case ?

4. Logic linking data to propositions
+ 5. Criteria for interpreting findings
- Chain of evidence from multiple sources
- When does data confirm proposition ? When does it refute ?

Threats to validity

---------Low hanging fruit---------

Research questions for Case Studies

Existence: ***Exploratory***
- Does X exist?

Description & Classification
- What is X like?
- What are its properties?
- How can it be categorized?
- How can we measure it?
- What are its components?

Descriptive-Comparative
- How does X differ from Y?

Frequency and Distribution
- How often does X occur?
- What is an average amount of X?

Descriptive-Process
- How does X normally work?
- By what process does X happen?
- What are the steps as X evolves?

Relationship ***Explanatory***
- Are X and Y related?
- Do occurrences of X correlate with occurrences of Y?

Causality
- What causes X?
- What effect does X have on Y?
- Does X cause Y?
- Does X prevent Y?

Causality-Comparative
- Does X cause more Y than does Z?
- Is X better at preventing Y than is Z?
- Does X cause more Y than does Z under one condition but not others?

Design
- What is an effective way to achieve X?
- How can we improve X?

Source: Empirical Research Methods in Requirements Engineering.
Tutorial given at RE'07, New Delhi, India, Oct 2007.

Units of Analysis

What phenomena to analyze
- depends on research questions
- affects data collection & interpretation
- affects generalizability

Possibilities
- individual developer
- a team
- a decision
- a process
- a programming language
- a tool

Example: Clone Detection, Bug Prediction
- the tool/algorithm
  Does it work ?
- the individual developer
  How/why does he produce bugs/clones ?
- about the culture/process in the team
  How does the team prevent bugs/clones ?
  How successful is this prevention ?
- about the programming language
  How vulnerable is the programming language towards clones / bugs ?
  (COBOL vs. AspectJ)

Design in advance
- avoid "easy" units of analysis
  + cases restricted to Java because parser
    - Is the language really an issue for your research question ?
  + report size of the system (KLOC, # Classes, # Bug reports)
    - Is team composition not more important ?

Threats to validity

- Source: Runeson, P. and Höst, M. 2009. Guidelines for conducting and reporting case study research in software engineering.

1. Construct validity
- Do the operational measures reflect what the researcher had in mind ?

2. Internal validity
- Are there any other factors that may affect the case ?
  ➡ Mainly when investigating causality !

3. External validity
- To what extent can the findings be generalized ?
  ➡ Precise research question & units of analysis required

4. Reliability
- To what extent is the data and the analysis dependent on the researcher (the instruments, …)

Other categories have been proposed as well
- credibility, transferability, dependability, confirmability

Threats to validity (Examples)

1. Construct validity
- Time recorded vs. time spent
- Execution time, memory consumption, …
  + noise of operating system, sampling method
- Human-assigned classifiers (bug severity, …)
  + risk for "default" values
- Participants in interviews have pressure to answer positively

2. Internal validity
- Were phenomena observed under special conditions
  + in the lab, close to a deadline, company risked bankruptcy, …
  + major turnover in team, contributors changed (open-source), …
- Similar observations repeated over time (learning effects)

3. External validity
- Does it apply to other languages ? other sizes ? other domains ?
- Background & education of participants
- Simplicity & scale of the team
  + small teams & flexible roles vs. large organizations & fixed roles

4. Reliability
- Bugs in the tool, the instrument ? Appropriate metrics & statistics ?
- Classification: if others were to classify, would they obtain the same ?