

Lecture 2: Science, SE Research & Empirical Studies

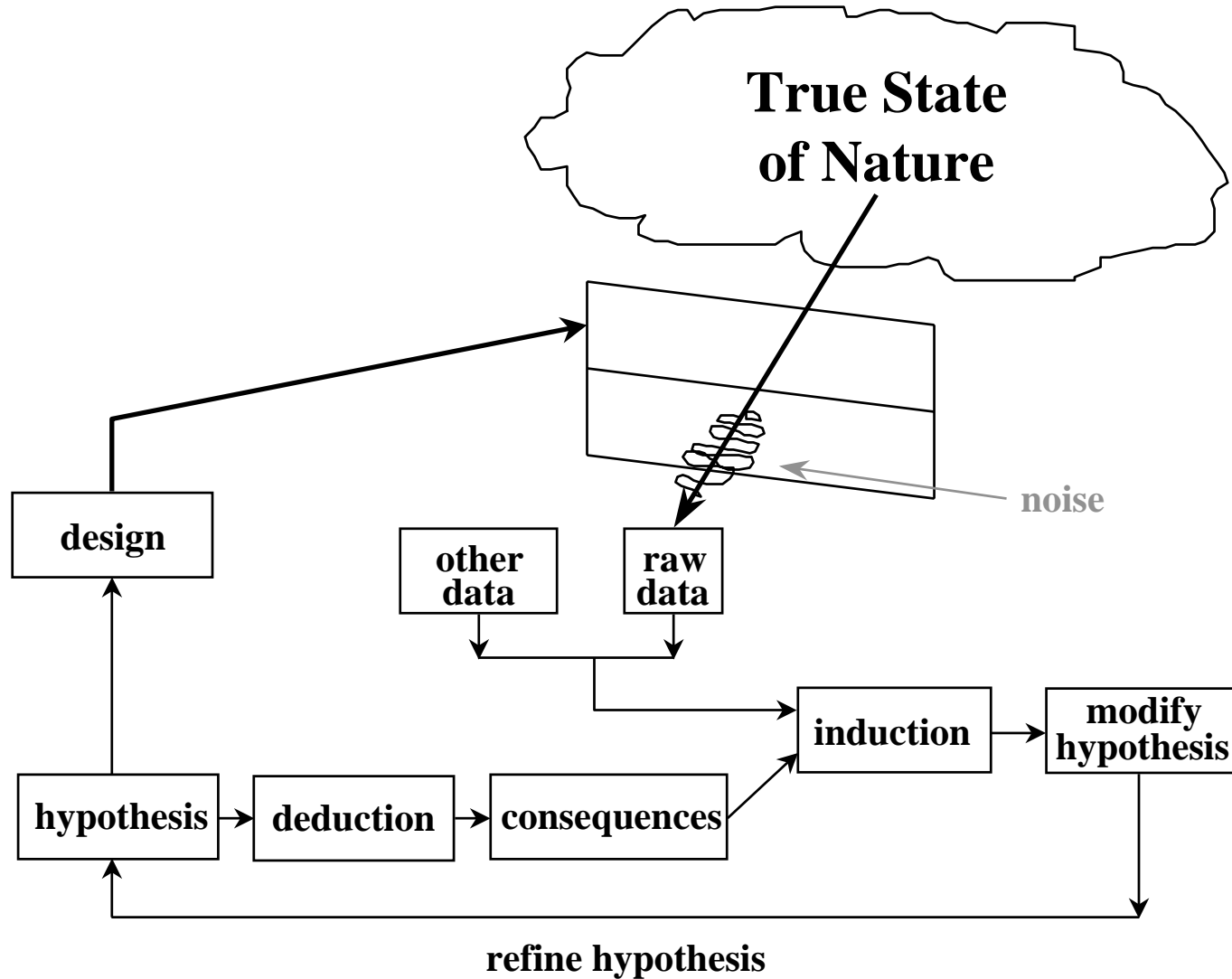
Dewayne E Perry

ENS 623

perry@ece.utexas.edu

[adapted in part from Steve Easterbrook, U Toronto]

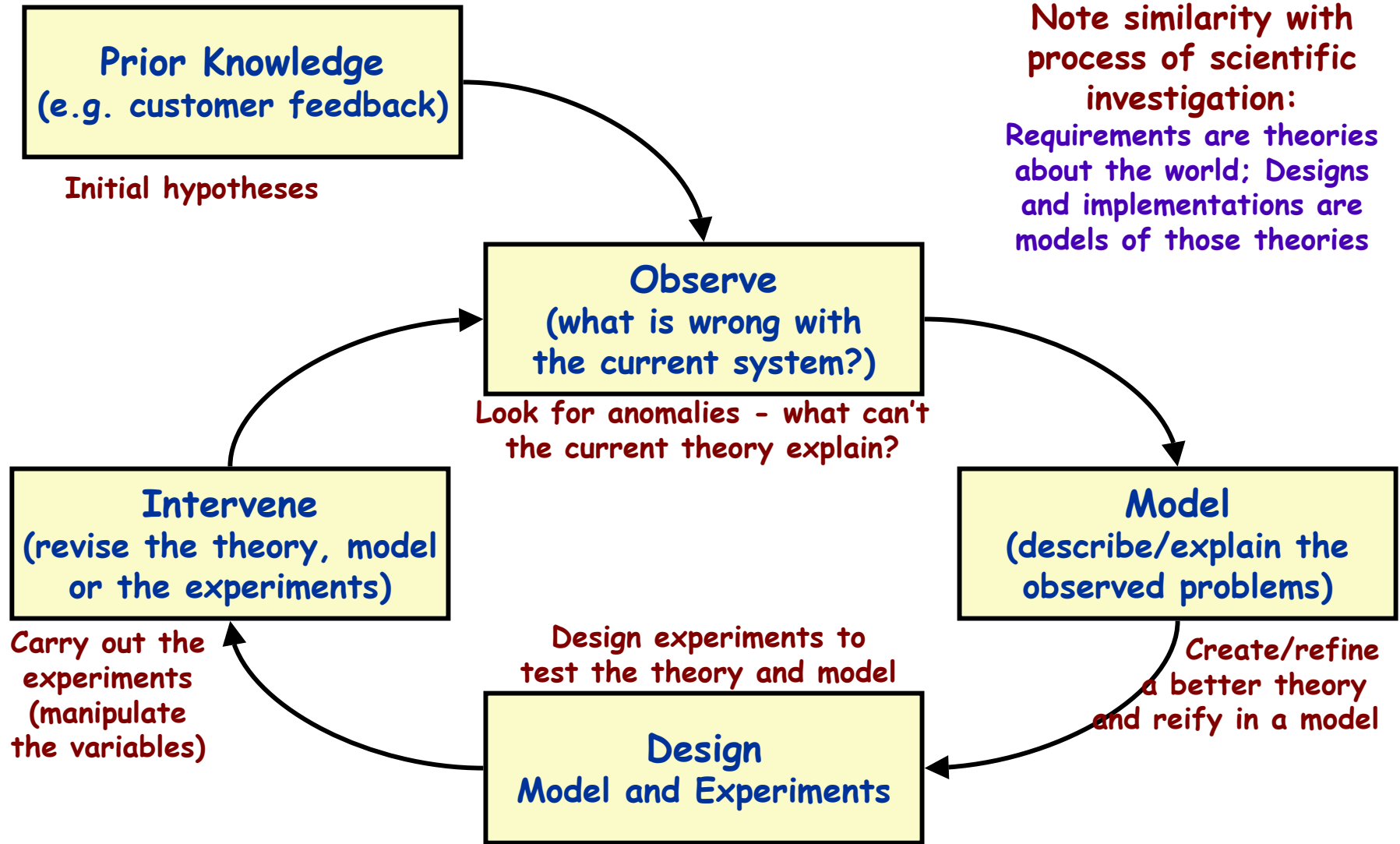
Reconciling Theory with Reality



Definitions

- ➔ An empirical study is a study reconciling theory and reality.
- ➔ Anecdotal and case studies are empirical studies that investigate phenomena of a current theory in its real-life context.
- ➔ An experiment is an empirical study that shows a mechanism by directly manipulating the independent factors to elicit a dependent factors' predicted (from theory) responses.

Inquiry Cycle



What is Science?

➔ Logical Positivism:

- ↪ Separates discovery from validation
- ↪ Logical deduction, to link theoretical concepts to observable phenomena
- ↪ Scientific truth is absolute, cumulative, and unifiable

➔ Popper:

- ↪ Theories can be refuted, not proved;
- ↪ Only falsifiable theories are scientific

➔ Campbell:

- ↪ Theories are underdetermined;
- ↪ All observation is theory-laden & biased

➔ Quine:

- ↪ Terms used in scientific theories have contingent meanings
- ↪ Cannot separate theoretical terms from empirical findings

➔ Kuhn:

- ↪ Science characterized by dominant paradigms, punctuated by revolution

What is Science?

➔ Lakatos:

- ↳ Not one paradigm, but many competing research programs
- ↳ Each has a hard core of assumptions immune to refutation

➔ Feyerabend:

- ↳ Cannot separate scientific discovery from its historical context
- ↳ All scientific methods are limited;
- ↳ Any method offering new insight is okay

➔ Toulmin:

- ↳ Evolving Weltanschauung determines what is counted as fact;
- ↳ Scientific theories describe ideals, and explain deviations

➔ Laudan:

- ↳ Negative evidence is not so significant in evaluating theories.
- ↳ All theories have empirical difficulties
- ↳ New theories seldom explain everything the previous theory did

Postmodernism and Science

➔ Modernism

- ↳ Rationality is the highest form of mental functioning
- ↳ Modern science produces universal truths
 - ...independent from the context and status of the scientist who produced them
- ↳ Rationality will always lead to progress and perfection
 - All human institutions can be scientifically analyzed and improved
- ↳ Reason is the ultimate judge of what is right (true, legal, ethical,...)
- ↳ Language must be rational
 - It only exists to represent the real world;
 - There must be a firm, objective connection between the "signifier" and the "signified"
 - The meaning cannot depend on the audience

Postmodernism and Science

➔ Postmodernism

↳ Questioning the grand narrative

- A grand narrative is a story that a culture/society tells itself about its practices and beliefs
- E.g. in the US: "democracy is the most enlightened/rational form of government"
- E.g. in science: "scientific truths are universal and eternal"
- Postmodernism identifies and critiques such narratives

↳ Instead, look for mini-narratives

- Stories that explain small practices, local events, situated, contingent behavior
- ...and don't make any claims about universality, truth, or stability

↳ E.g. Literary Deconstruction

- Examine what a text does not say, what it represses
- Reveal internal arbitrary hierarchies and dichotomies

↳ E.g. Semiotics

- The study of the relationship between signs and the things they signify

What is Engineering?

➔ Traditional View:

Scientists...

create knowledge
study the world as it is
are trained in scientific method
design
use explicit knowledge
are thinkers

Engineers...

apply that knowledge
seek to change the world
are trained in engineering

use tacit knowledge
are doers

➔ More realistic View

Scientists...

create knowledge
are problem-driven
seek to understand and explain
design experiments to test theories
prefer abstract knowledge
but rely on tacit knowledge

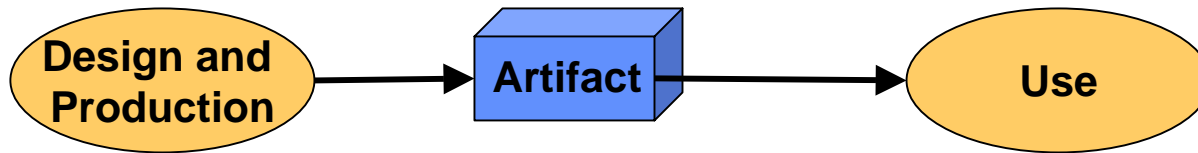
Engineers...

create knowledge
are problem-driven
seek to understand and explain
design devices to test theories
prefer contingent knowledge
but rely on tacit knowledge

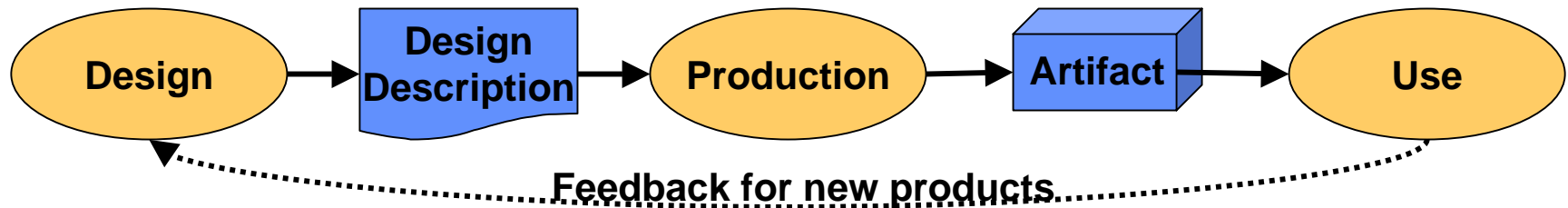
Both involve a mix of design and discovery

What is (Software) Design?

Pre-industrial design:



Industrial design:



Software Design:



Normal vs Radical design

➔ Normal design:

↪ Old problems, whose solutions are well known

- Engineering codifies standard solutions
- Engineer selects appropriate methods and technologies

↪ Design focuses on well understood devices

- Devices can be studied independent of context
- Differences between the mathematical model and the reality are minimal

➔ Radical design:

↪ Never been done, or past solutions have failed

- Often the challenge is to deal with a very complex problem

↪ Bring together complex assemblies of devices into new systems

- Such systems are not amenable to reductionist theories
- Such systems are often soft: no objective criteria for describing the system

➔ Examples:

↪ Most of Electrical Engineering involves normal design

↪ All of Systems Engineering involves radical design (by definition!)

↪ The part of S/W Eng concerned with human activities is radical design - what else?

SE Weltanschauung

- ➔ **Software-intensive systems**
 - ↪ Software + hardware + human activity
 - the human activity gives a system its purpose
 - ↪ RE is about discovering that purpose
 - ↪ SE is about satisfying that purpose
- ➔ **Continuous Change**
 - ↪ Introduction of new system changes the human activity
 - ↪ People find new ways of using it
- ➔ **Human Centered Development**
 - ↪ Goal is to change human activities...
 - ...to make them more effective, efficient, safe, enjoyable, etc.
 - ↪ ...rather than to design a new computer system
- ➔ **A Systems Perspective**
 - ↪ Treat relevant parts of the world as systems with emergent properties
- ➔ **Multi-disciplinary approach**
 - ↪ Use whatever techniques seem useful
 - Social, cognitive, mathematical,...
- ➔ **Design as Reflection**
 - ↪ New designs arise in response to observed problems with existing ones
 - ↪ There is always an existing system!
- ➔ **Form does not follow function**
 - ↪ Because we only understand the function properly in hindsight
- ➔ **Multiple Viewpoints**
 - ↪ Many stakeholders
 - ↪ Each model presupposes a viewpoint
 - ↪ All models are imperfect
- ➔ **Negotiation is central**
 - ↪ Resolve conflicts between different stakeholders' goals
 - ↪ Manage customer's expectations

Many Methods Available:

- ➔ **Laboratory Experiments**
- ➔ **Field Studies**
- ➔ **Case Studies**
- ➔ **Pilot Studies**
- ➔ **Rational Reconstructions**
- ➔ **Exemplars**
- ➔ **Surveys**
- ➔ **Artifact/Archive Analysis (“mining”!)**
- ➔ **Ethnographies**
- ➔ **Action Research**
- ➔ **Simulations**
- ➔ **Benchmarks**

Laboratory Experiments

Experimental investigation of a testable hypothesis, in which conditions are set up to isolate the variables of interest ("independent variables") and test how they affect certain measurable outcomes (the "dependent variables")

➔ Good for

- ↪ Quantitative analysis of benefits of a particular tool/technique
- ↪ (Demonstrating how scientific we are!)

➔ Limitations

- ↪ Hard to apply if you cannot simulate the right conditions in the lab
- ↪ Limited confidence that the lab setup reflects the real situation
- ↪ Ignores contextual factors (e.g. social/org'al/political factors)
- ↪ Extremely time-consuming!

See:

- Pfleeger, S.L.; Experimental design and analysis in software engineering. *Annals of Software Engineering* 1, 219-253. 1995
- D. Perry, A. Porter, L. Votta "Empirical Studies of Software Engineering: A Roadmap". In A. Finkelstein (ed) "The Future of Software Engineering". IEEE CS Press, 2000.

Field Studies

Exploratory study, used where little is currently known about a problem, or where we wish to check that our research goals are grounded in real-life settings; studies organizational practice using anthropological techniques.

➔ Good for

- ↪ Setting a research agenda (what really matters?)
- ↪ Understanding the context for RE problems (naturalistic inquiry)

➔ Limitations

- ↪ Hard to build generalizations (results may be organization specific)
- ↪ Observers' bias

See:

Klein, H.K., and Myers, M.D. "A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems," *MIS Quarterly*, vol 23 No 1, pp67-93, 1999.

Case Studies

A technique for detailed exploratory investigations, both prospectively and retrospectively, that attempt to understand and explain phenomenon or test theories, using primarily qualitative analysis

➔ Good for

- ↪ Answering detailed how and why questions
- ↪ Gaining deep insights into chains of cause and effect
- ↪ Testing theories in complex settings where there is little control over the variables

➔ Limitations

- ↪ Hard to find appropriate case studies
- ↪ Hard to quantify findings

See:

Flyvbjerg, B.; Five Misunderstandings about Case Study Research. *Qualitative Inquiry* 12 (2) 219-245, April 2006

Pilot Studies

Controlled introduction of a tool/technique into a real project, where the researcher can no longer control the context, but where the net effect can be measured (e.g. against a baseline, or against previous experience)

→ Good for

- ↳ Measuring the benefits in a real setting
- ↳ Preparation for tech. transfer
- ↳ Getting organizations interested in your work

→ Limitations

- ↳ Hard to get organizations to adopt unproven ideas
- ↳ Hawthorn effect (and other bias problems)

See:

R. L Glass "Pilot Studies: What, Why and How" J. Systems and Software, vol 36, no 1, pp85-97, 1997

Rational Reconstructions

A demonstration of a tool or technique on data taken from a real case study, but applied after the fact to demonstrate how the tool/technique would have worked

➔ Good for

- ↳ Initial validation before expensive pilot studies
- ↳ Checking the researcher's intuitions about what the tool/technique can do

➔ Limitations

- ↳ potential bias (you knew the findings before you started)
- ↳ easy to ignore "signal-to-noise ratio"

➔ Examples

- ↳ LAS; BART; ... etc.

See:

Examples in Cohen *Empirical Methods for Artificial Intelligence*

Exemplars

Self-contained, informal descriptions of a problem in some application domain; exemplars are to be considered immutable; the specifier must do the best she can to produce a specification from the problem statement.

➔ **Good for:**

- ↪ Setting research goals,
- ↪ Understanding differences between research programs

➔ **Limitations:**

- ↪ No clear criteria for comparing approaches
- ↪ Not clear that "immutability" is respected in practice

➔ **Examples:**

- ↪ Meeting Scheduler; Library System; Elevator Control System; Telephones;...

See:

M. S. Feather, S. Fickas, A. Finkelstein, and A. van Lamsweerde, "Requirements and Specification Exemplars," *Automated Software Engineering*, vol. 4, pp. 419-438, 1997.

Surveys

A comprehensive system for collecting information to describe, compare or explain knowledge, attitudes and behaviour over large populations

➔ Good for

- ↪ Investigating the nature of a large population
- ↪ Testing theories where there is little control over the variables

➔ Limitations

- ↪ Relies on self-reported observations
- ↪ Difficulties of sampling and self-selection
- ↪ Information collected tends to subjective opinion

See:

Shari Lawrence Pfleeger and Barbara A. Kitchenham,
"Principles of Survey Research," Software Engineering
Notes, (6 parts) Nov 2001 - Mar 2003

Artifact / Archive Analysis

Investigation of the artifacts (documentation, communication logs, etc) of a software development project after the fact, to identify patterns in the behaviour of the development team.

→ Good for

- ↳ Understanding what really happens in software projects
- ↳ Identifying problems for further research

→ Limitations

- ↳ Hard to build generalizations (results may be project specific)
- ↳ Incomplete data

See:

Audris Mockus, Roy T. Fielding, and James Herbsleb. Two case studies of open source software development: Apache and mozilla. *ACM Transactions on Software Engineering and Methodology*, 11(3):1-38, July 2002.

Ethnographies

Interpretive, in-depth studies in which the researcher immerses herself in a social group under study to understand phenomena through the meanings that people assign to them

➔ Good for:

- ↪ Understanding the intertwining of context and meaning
- ↪ Explaining cultures and practices around tool use

➔ Limitations:

- ↪ No generalization, as context is critical
- ↪ Little support for theory building

See:

Klein, H. K.; Myers, M. D.; A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems. MIS Quarterly 23(1) 67-93. March 1999.

Action Research

Research and practice intertwine and shape one another. The researcher mixes research and intervention and involves organizational members as participants in and shapers of the research objectives

➔ Good for

- ↪ Any domain where you cannot isolate variables, cause from effect...
- ↪ Ensuring research goals are relevant
- ↪ When effecting a change is as important as discovering new knowledge

➔ Limitations

- ↪ Hard to build generalizations (abstractionism vs. contextualism)
- ↪ Won't satisfy the positivists!

See:

- Lau, F; Towards a framework for action research in information systems studies. *Info.Technology and People* 12 (2) 148-75. 1999.
- Kock, N.F., (1997), Myths in Organisational Action Research: Reflections on a Study of Computer-Supported Process Redesign Groups, *Organizations & Society*, V.4, No.9, pp. 65-91.

Simulations

An executable model of the software development process, developed from detailed data collected from past projects, used to test the effect of process innovations

➔ Good for:

- ↳ Preliminary test of new approaches without risk of project failure
- ↳ [Once the model is built] each test is relatively cheap

➔ Limitations:

- ↳ Expensive to build and validate the simulation model
- ↳ Model is only as good as the data used to build it
- ↳ Hard to assess scope of applicability of the simulation

See:

Kellner, M. I.; Madachy, R. J.; Raffo, D. M.; Software Process Simulation Modeling: Why? What? How? Journal of Systems and Software 46 (2-3) 91-105, April 1999.

Benchmarks

A test or set of tests used to compare alternative tools or techniques. A benchmark comprises a motivating comparison, a task sample, and a set of performance measures

➔ Good for

- ↪ making detailed comparisons between methods/tools
- ↪ increasing the (scientific) maturity of a research community
- ↪ building consensus over the valid problems and approaches to them

➔ Limitations

- ↪ can only be applied if the community is ready
- ↪ become less useful / redundant as the research paradigm evolves

See:

S. Sim, S. M. Easterbrook and R. C. Holt "Using Benchmarking to Advance Research: A Challenge to Software Engineering". Proceedings, ICSE-2003

Questions

- ➔ Do any of these idioms capture your research?
 - ↳ Do the distinctions make sense?
 - ↳ Are there other idioms we've missed?
- ➔ Are we (as a community) using the right idioms?
 - ↳ Should we be using some of them more than we do?
 - ↳ Should we be using some of them less than we do?
- ➔ What standards of reporting should we demand?
 - ↳ Eg, when reviewing papers for SE conferences
 - ↳ Should we be more explicit about our research methods?
- ➔ What practical steps can we take...
 - ↳ Workshops on research validation?
 - ↳ Benchmarking initiatives?

Validating SE models

➔ Logical Positivist view:

- "There is an objective world that can be modeled by building a consistent body of knowledge grounded in empirical observation"

↪ In SE: "there is an objective problem that exists in the world"

- Build a consistent theory/model; make sufficient empirical observations to check validity
- Use tools that test consistency and completeness of the theory/model
- Use reviews, prototyping, etc to demonstrate the theory/model is "valid"

➔ Popper's modification to logical positivism:

- "Theories can't be proven correct, they can only be refuted by finding exceptions"

↪ In SE: "theories and models must be refutable"

- Look for evidence that the model is wrong
- Eg, collect scenarios and check the model supports them

➔ Post-Modern view:

- "There is no privileged viewpoint; all observation is value-laden; scientific investigation is culturally embedded"

- Eg, Kuhnian paradigms; Toulmin's *weltanschauungen*

↪ In SE: "validation is always subjective and contextualised"

- Use stakeholder involvement so that they 'own' the requirements
- Use ethnographic techniques to understand the *weltanschauungen*