# CSC2130: Empirical Research Methods for Computer Scientists

**Steve Easterbrook** sme@cs.toronto.edu

**Barbara Neves** barbara@bbneves.com

**www.cs.toronto.edu/~sme/CSC2130/**







## Course Goals

- ➜ **Motivate the need for an empirical basis for research claims**
- ➜ **Prepare students for advanced research:**
  - Learn how to plan, conduct and report on empirical investigations.
  - Understand the key steps of a research project:
    - formulating research questions,
    - theory building,
    - data analysis (using both qualitative and quantitative methods),
    - building evidence,
    - assessing validity,
    - publishing.
- ➜ **Cover the principal empirical methods applicable to human subjects studies in CS**
  - controlled experiment, case studies, surveys, archival analysis, action research, ethnographies,…
- ➜ **Relate these methods to relevant meta-theories in the philosophy and sociology of science.**

# Intended Audience

- **This is an advanced course:**
  - assumes a strong grasp of the key research questions in your own research area, and that you are already doing independent research
- **Focus:**
  - How do people use computer technology?
  - How does this technology (re-)shape human activities?
  - How can we apply qualitative and quantitative techniques from the behavioural sciences to help answer these questions?
- **The course is aimed at students who:**
  - …plan to conduct research (in SE, HCI, etc) that demands some empirical validation
  - …wish to establish an empirical basis for an existing research programme
  - …wish to apply these techniques in related fields (e.g. Cog Sci, )
- **Note: we will *not* cover the kinds of experimental techniques used in CS systems areas, nor in medical/biological research**
  - Focus is on the relationship between human activity and computer technology






# Format

- **Seminars:**
  - 1 three-hour seminar per week
  - Mix of discussion, lecture, student presentations
- **Readings**
  - Major component is discussion of weekly readings
  - Please read the set papers before the seminar
- **Assessment:**
  - 10% Class Participation
  - 20% Oral Presentation – introduce the readings using no more than two slides
    - Slide 1: Key takeaway messages from the paper
    - Slide 2: Discussion questions
  - 70% Written paper
    - A critical literature review and study design for a specific research question
    - preferably related to your own research

# Course Outline (part 1)

1. Introduction & Orientation (today!)
   - Intro to philosophy & sociology of science
   - Role of theory building
2. Research Design and Ethics
   - What counts as evidence?
   - Use of mixed methods
   - Ethics concerns for human subjects studies
3. Basics of Doing Research
   - Finding good research questions
   - Theory building
   - Evidence and Measurement
   - Replication
   - Peer Review






# Course Outline (cont)

4. Laboratory Experiments
   - Controlled Experiments
   - Quasi-experiments
   - Sampling
   - Confounding Variables
5. Quantitative Analysis
   - Basic Stats
   - Interpreting significance measures
   - power analysis
6. Interviews and Observation
   - Conducting Interviews
   - Participant observation
   - Collecting field notes
7. Qualitative Analysis
   - Coding Strategies
   - Grounded Theory
   - Phenomenography
8. Case Studies
   - Single and Multi-case
   - Longitudinal Case Studies
9. Survey Research
   - Questionnaire Design
   - Sample Size
10. Intervention Methods
    - Action Research
    - Pilot Studies
    - Benchmarking
11. Publishing and Reviewing
    - (mock PC meeting)
12. Replication and Beyond
    - Internal and External Replication
    - Biases and Influences
    - Threats to Validity
    - When to use empirical methods

# Is this your research plan?

**Step 1: Build a new tool**

**Step 2: ??**

**Step 3: Profit**






# Engineering vs. Science

- **Traditional View:**

| Scientists… | Engineers… |
|---|---|
| create knowledge | apply that knowledge |
| study the world as it is | seek to change the world |
| are trained in scientific method | are trained in engineering design |
| use explicit knowledge | use tacit knowledge |
| are thinkers | are doers |

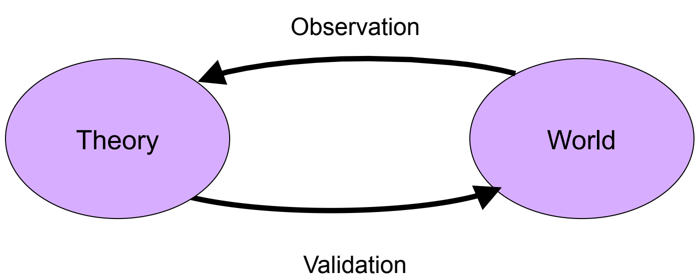
- **More realistic View**

| Scientists… | Engineers… |
|---|---|
| create knowledge | create knowledge |
| are problem-driven | are problem-driven |
| seek to understand and explain | seek to understand and explain |
| design experiments to test theories | design devices to test theories |
| prefer abstract knowledge | prefer contingent knowledge |
| but rely on tacit knowledge | but rely on tacit knowledge |

**Both involve a mix of design and discovery**

# Scientific Method

- No single "official" scientific method
- Somehow, scientists are supposed to do this:



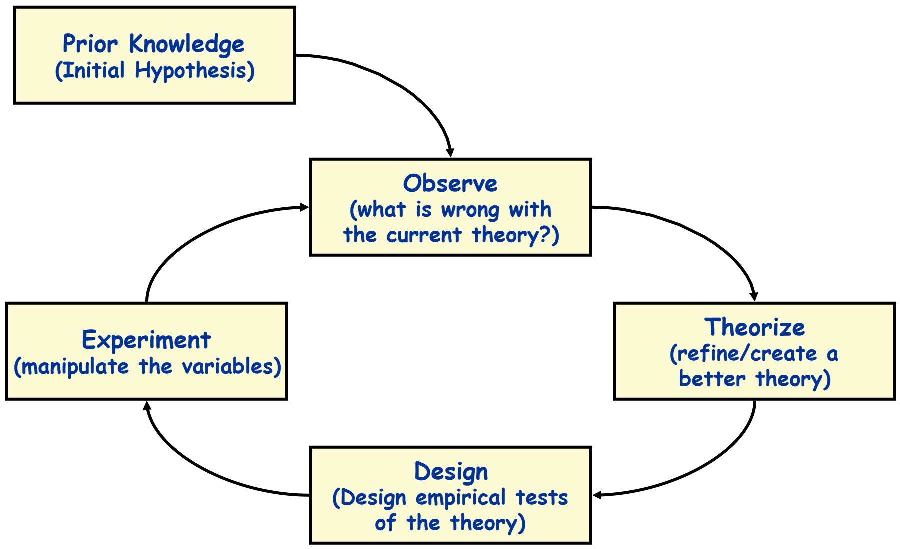






Observe!

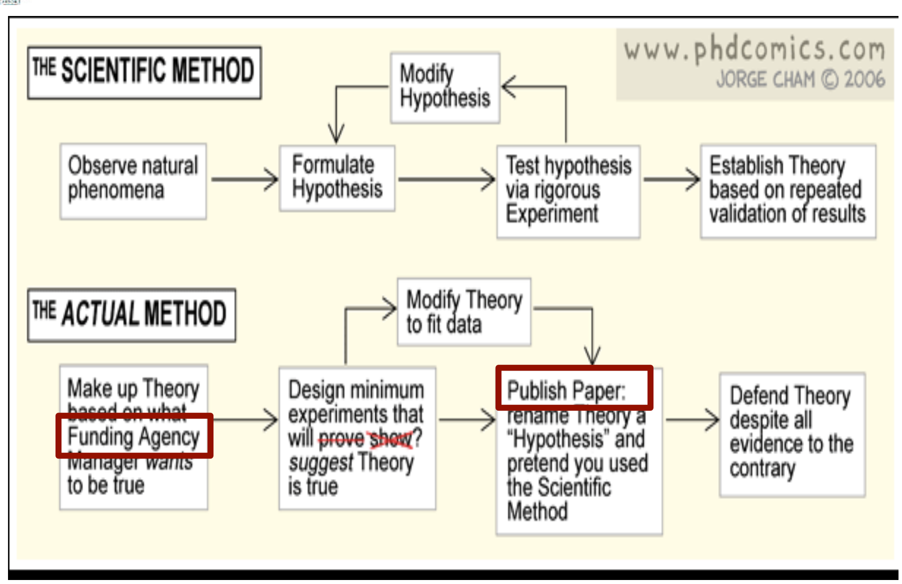
## Scientific Inquiry











# Some Characteristics of Science

- Science seeks to improve our understanding of the world.
- Explanations are based on observations
  - Scientific truths must stand up to empirical scrutiny
  - Sometimes "scientific truth" must be thrown out in the face of new findings
- Theory and observation affect one another:
  - Our perceptions of the world affect how we understand it
  - Our understanding of the world affects how we perceive it
- Creativity is important
  - Theories, hypotheses, experimental designs
  - Search for elegance, simplicity






# All Methods are flawed

- E.g. Laboratory Experiments
  - Cannot study large scale software development in the lab!
  - Too many variables to control them all!
- E.g. Case Studies
  - How do we know what's true in one project generalizes to others?
  - Researcher chose what questions to ask, hence biased the study
- E.g. Surveys
  - Self-selection of respondents biases the study
  - Respondents tell you what they think they ought to do, not what they actually do
- ...etc...

# Strategies to overcome weaknesses

- **Theory-building**
  - Testing a hypothesis is pointless (single flawed study!)…
  - …unless it builds evidence for a clearly stated theory
- **Empirical Induction**
  - Series of studies over time…
  - Each designed to probe more aspects of the theory
  - …together build evidence for a clearly stated theory
- **Mixed Methods Research**
  - Use multiple methods to investigate the same research question
  - Each method compensates for the flaws of the others
  - …together build evidence for a clearly stated theory







Source: http://xkcd.com/882/

# What is a research contribution?

- → A better understanding of how people use software technology?
- → Identification of problems with the current state-of-the-art?
- → A characterization of the properties of new tools/techniques?
- → Evidence that approach A is better than approach B?

**How will you validate your claims?**






# Meet Stuart Dent

- → **Name:**
  - Stuart Dent (a.k.a. "Stu")
- → **Advisor:**
  - Prof. Helen Back
- → **Topic:**
  - Merging Stakeholder views in Model Driven Development
- → **Status:**
  - 2 years into his PhD
  - Has built a tool
  - Needs an evaluation plan

# Stu's Evaluation Plan

→ **Formal Experiment**
- Independent Variable: Stu-Merge vs. Rational Architect
- Dependent Variables: Correctness, Speed, Subjective Assessment
- Task: Merging Class Diagrams from two different stakeholders' models
- Subjects: Grad Students in SE
- $H_1$: "Stu-Merge produces correct merges more often than RA"
- $H_2$: "Subjects produce merges faster with Stu-Merge than with RA"
- $H_3$: "Subjects prefer using Stu-Merge to RA"

→ **Results**
- $H_1$ accepted (strong evidence)
- $H_2$ & $H_3$ rejected
- Subjects found the tool unintuitive






# Threats to Validity

→ **Construct Validity**
- What do we mean by a merge? What is correctness?
- 5-point scale for subjective assessment - insufficient discriminatory power
  - (both tools scored very low)

→ **Internal Validity**
- Confounding variables: Time taken to learn the tool; familiarity
  - Subjects were all familiar with RA, not with Stu-merge

→ **External Validity**
- Task representativeness
  - class models were of a toy problem
- Subject representativeness
  - Grad students as sample of what population?

→ **Theoretical Reliability**
- Researcher bias
  - subjects knew Stu-merge was Stu's own tool

## What went wrong?

- What was the research question?
  - "Is tool A better than tool B?"
- What would count as an answer?
- What use would the answer be?
  - How is it a "contribution to knowledge"?
- How does this evaluation relate to the existing literature?






## Experiments as Clinical Trials

Why would we expect it to be better?

Why do we need to know?

What will we do with the answer?

**Is drug A better than drug B?**

Better at doing what?

Better in what way?

Better in what situations?

Why would we expect it to be better?

You gotta have a theory!






## Some Definitions

- A **model** is an abstract representation of a phenomenon or set of related phenomena
  - Some details included, others excluded
- A **theory** is a set of statements that explain a set of phenomena
  - Serves to explain and predict
  - Precisely defined terminology
  - Concepts, relationships, causal inferences
  - (operational definitions for theoretical terms)
- A **hypothesis** is a testable statement derived from a theory
  - A hypothesis is not a theory!
- In CS, we have mostly *folk theories*

## A simpler definition

A (good) Theory is the best explanation of all the available evidence






## The Role of Theory Building

- Theories lie at the heart of what it means to do science.
  - Production of generalizable knowledge
- Theory provides orientation for data collection
  - Cannot observe the world without a theoretical perspective
- Theories allow us to compare similar work
  - Theories include precise definition for the key terms
  - Theories provide a rationale for which phenomena to measure
- Theories support analytical generalization
  - Provide a deeper understanding of our empirical results
  - …and hence how they apply more generally
  - Much more powerful than statistical generalization

# Stu's Theory

- **Background Assumptions**
  - Large team projects, models contributed by many actors
  - Models are fragmentary, capture partial views
  - Partial views are inconsistent and incomplete most of the time
- **Basic Theory**
  - (Brief summary:)
  - Model merging is an exploratory process, in which the aim is to discover intended relationships between views. 'Goodness' of a merge is a subjective judgment. If an attempted merge doesn't seem 'good', many need to change either the models, or the way in which they were mapped together.
  - [Still needs some work]
- **Derived Hypotheses**
  - Useful merge tools need to represent relationships explicitly
  - Useful merge tools need to be complete (work for any models, even if inconsistent)






# What type of question are you asking?

- **Existence:**
  - 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**
  - Are X and Y related?
  - Do occurrences of X correlate with occurrences of Y?
- **Causality**
  - Does X cause Y?
  - Does X prevent Y?
  - What causes X?
  - What effect does X have on 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?

## What type of question are you asking?

- **Existence:**
  - 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?

Exploratory

- **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?

Baserate

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

Correlation

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

Causal Relationship

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

Design



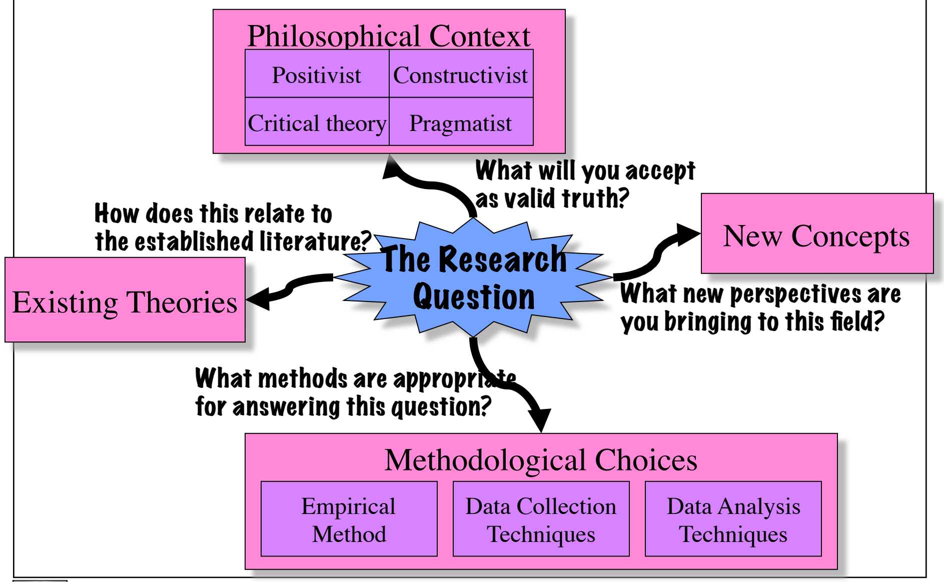



## Stu's Research Question(s)

- **Existence**
  - Does model merging ever happen in practice?
- **Description/Classification**
  - What are the different types of model merging that occur in practice on large scale systems?
- **Descriptive-Comparative**
  - How does model merging with explicit representation of relationships differ from model merging without such representations?
- **Causality**
  - Does an explicit representation of the relationship between models cause developers to explore different ways of merging models?
- **Causality-Comparative**
  - Does the algebraic representation of relationships in Stu's tool lead developers to explore more than do pointcuts in AOM?

Pick just one for now…

## Putting the Question in Context









## Many available methods…

**Common "in the lab" Methods**

- Controlled Experiments
- → Rational Reconstructions
- → Exemplars
- → Benchmarks
- → Simulations

**Common "in the wild" Methods**

- Quasi-Experiments
- Case Studies
- Survey Research
- Ethnographies
- Action Research

❍ Artifact/Archive Analysis ("mining"!)

# Stu's Method(s) Selection…

- Existence
  - Does model merging ever happen in practice?
- Description/Classification
  - What are the different types of model merging that occur in practice on large scale systems?
- Descriptive-Comparative
  - How does model merging with explicit representation of relationships differ from model merging without such representation?
- Causality
  - Does an explicit representation of the relationship between models cause developers to explore different ways of merging models?
- Causality-Comparative
  - Does the algebraic representation of relationships in Stu's tool lead developers to explore more than do pointcuts in AOM?

Case study

Survey Research

Ethnography

Action Research

Controlled Experiment






# Warning

**No method is perfect**

**Don't get hung up on methodological purity**

**Pick something and get on with it**

**Some knowledge is better than none**

University of Toronto Department of Computer Science

# Okay, but…

 35

University of Toronto Department of Computer Science

# Why Build a Tool?

- ➜ **Build a Tool to Test a Theory**
  - ✎ Tool is part of the experimental materials needed to conduct your study
- ➜ **Build a Tool to Develop a Theory**
  - ✎ Theory emerges as you explore the tool
- ➜ **Build a Tool to Explain your Theory**
  - ✎ Theory as a concrete instantiation of (some aspect of) the theory

Why did Stu build a tool?

?

 36