Conduct Rigorous and Scientific Research

Chang Xu

Department of Computer Science and Technology
Nanjing University
Oct 29, 2021

Outline

- I. Writing and Presentation
- II. Research Formulation
- III. Experimentation

III. Experimentation

1. Questions and subjects
2. Experimental design
3. Threats to validity
4. Exercise and discussion

General

- Experimentation
  - Is not merely a description of the experimental procedure and a list of experimental results
  - Should have a careful ***design*** (questions and variables) and discussion of potential ***threats*** (construct validity, internal validity, external validity, and conclusion validity)

Experimentation and Case Study

- Experimentation
  - In a lab environment
  - Variables (factors) can be isolated and ***controlled***
- Case study
  - Under an industrial (real-world) setting
  - Hard to repeat
- We mostly conduct ***controlled experiments***

Key Points (1)

- ***Questions*** to answer
  - Is Tool A ***better*** than Tool B?

Why would we expect it to be better?

Why do we need to know?

What will we do with the answer?

Better at doing what?

E.g., Study or physical exercise?

Better in what way?

E.g., Efficiency or durability?

Better in what situations?

E.g., In classroom or dorm?

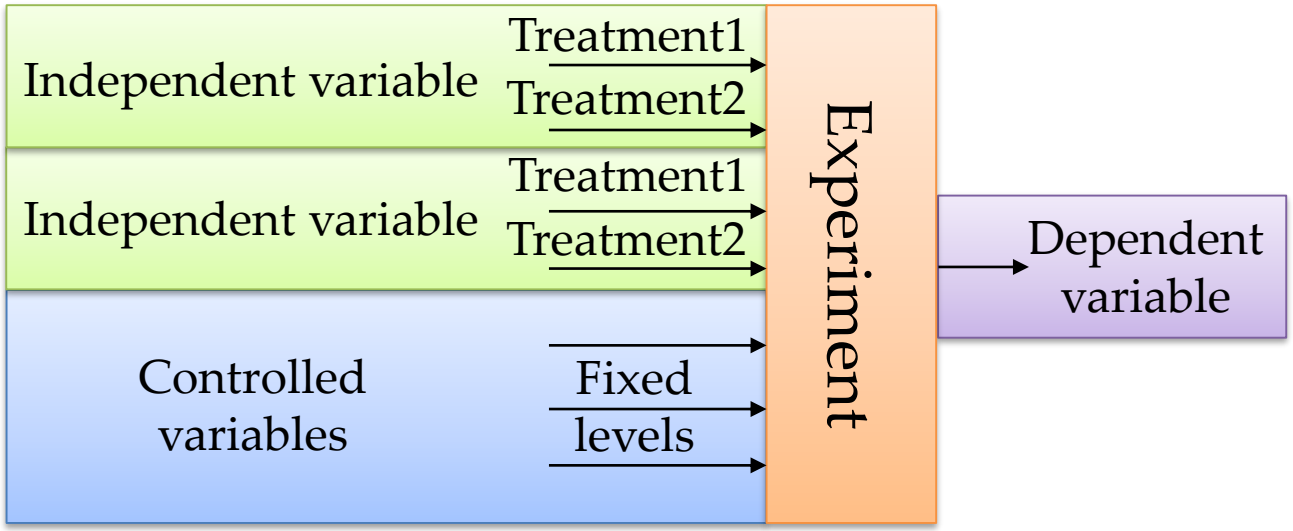
Key Points (2)

- ***Subjects*** selected
  - Sample of what population?
  - Consider the representativeness

- ***Variables*** and ***threats*** to validity
  - Variables: See the next page
  - Threats to validity: See an example

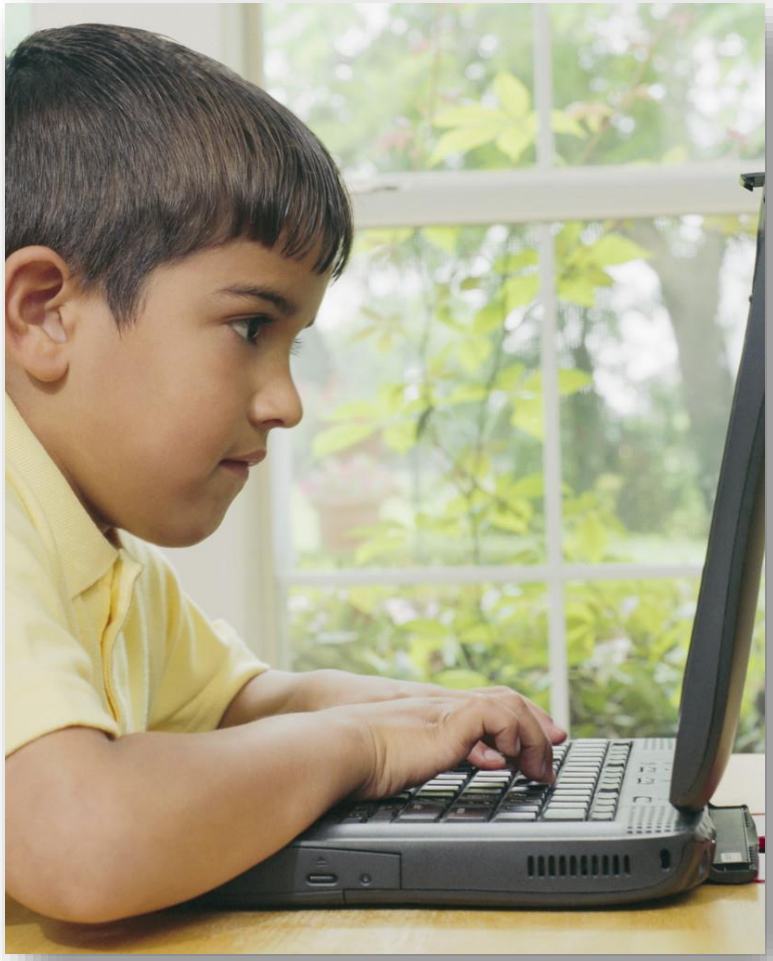
Variables

- Independent variables (factors)
- Dependent variables
- Controlled variables



Example

- Name
  - Stuart Bean ("stu")
- Topic
  - Merging stakeholder views in model-driven development
- Status
  - 2 years into his PhD study
  - Has built a tool
  - Needs evaluation

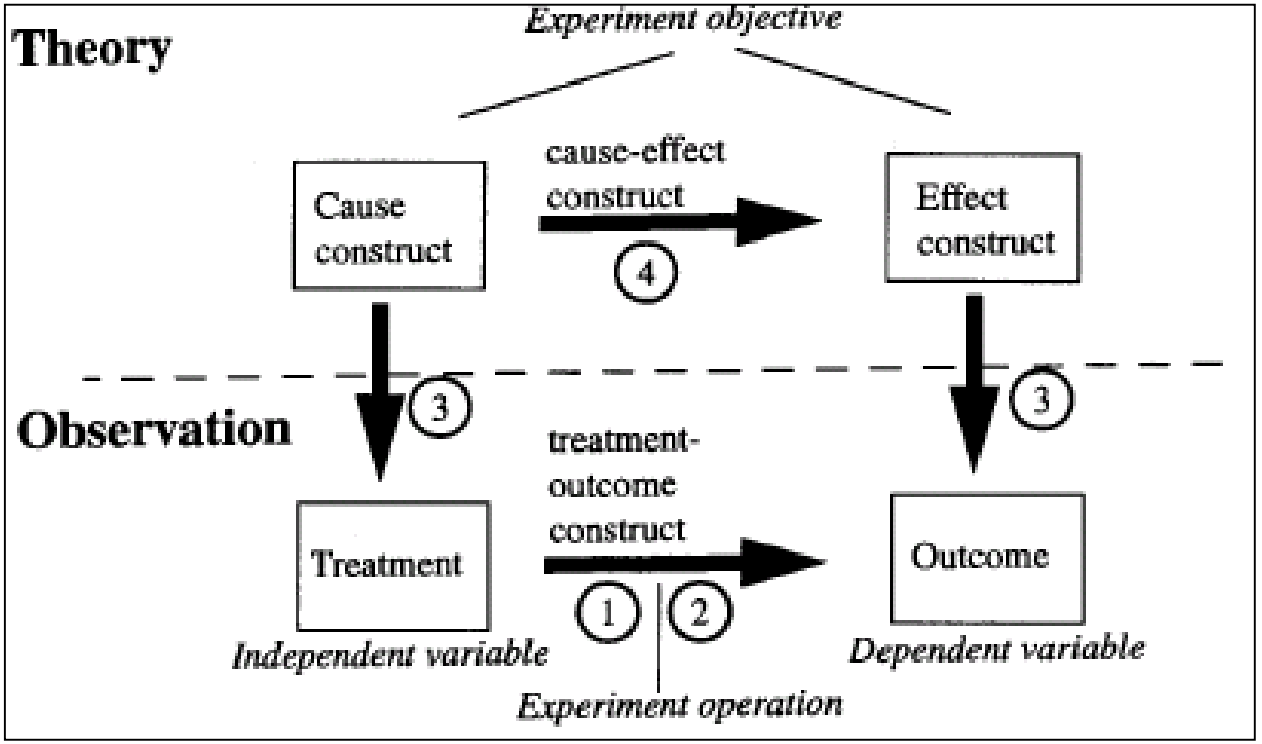


Stu's Evaluation Plan

- Experiments
  - Independent variable: Stu-merge vs. Rational Architect (RA)
  - Dependent variables: correctness, speed, assessment
  - Controlled variables: task (merging class diagrams from two different stakeholders' models), subjects (graduate students in software engineering)
- Hypotheses
  - H1: Stu-merge produces correct merges more often than RA
  - H2: Subjects produce merges faster with Stu-merge than RA
  - H3: Subjects prefer using Stu-merge to RA
- H1 accepted (strong evidence), H2 & H3 rejected

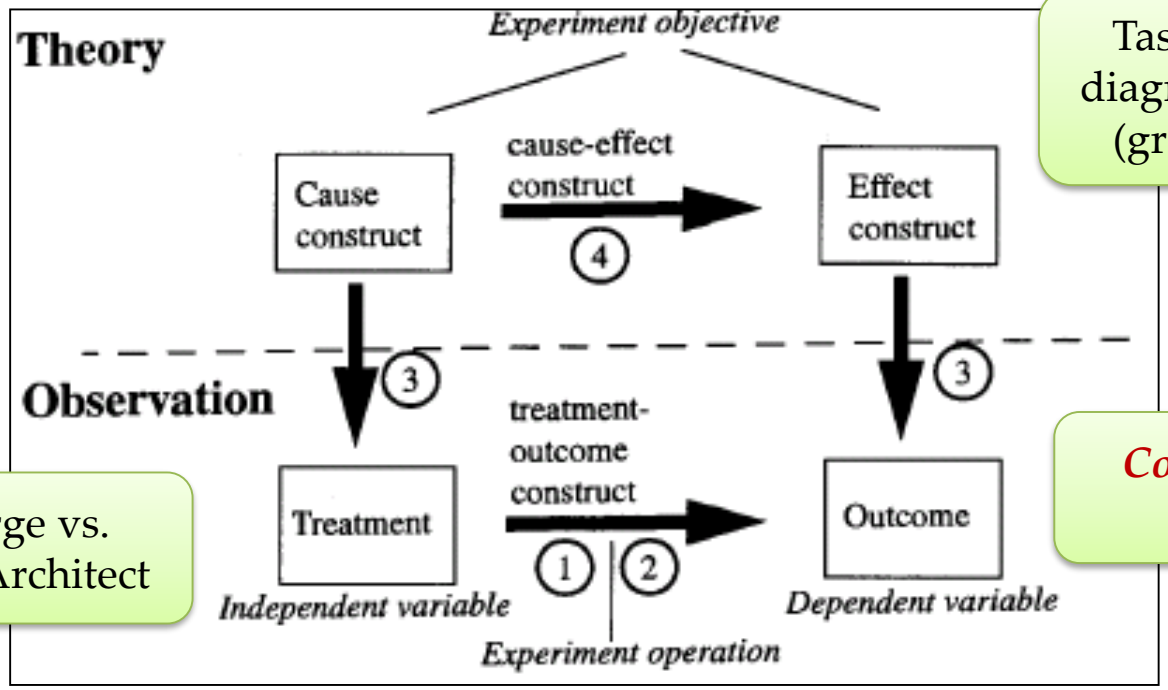
Threats to Validity Analysis

- ③ construct validity ② internal validity
- ④ external validity ① conclusion validity



Threats to Validity (1)

- ③ construct validity
  - What do we mean by a merge? What is correctness?
  - 0-5 point scale for subjective assessment - insufficient discriminatory power (both tools scored very low)



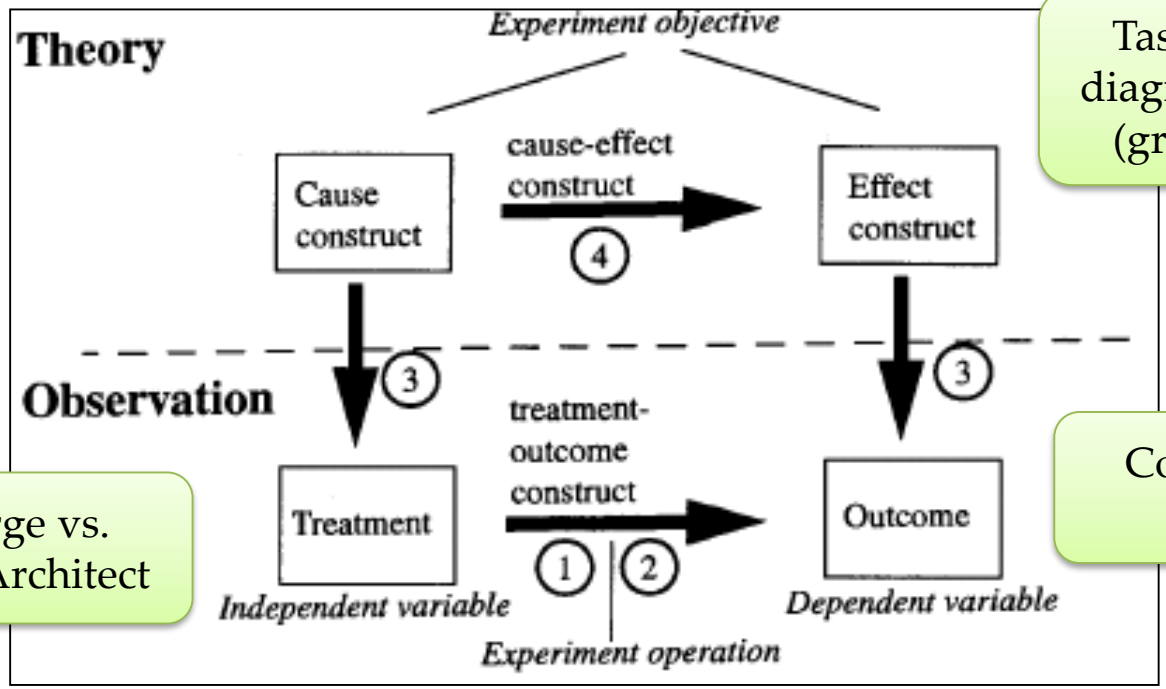
Task (merging class diagrams) and subjects (graduate students)

Stu-merge vs. Rational Architect

Correctness, speed, *assessment*

Threats to Validity (2)

- ② internal validity
  - ***Confounding*** variable: time taken to learn the tool (subjects were all familiar with RA, not with Stu-merge)



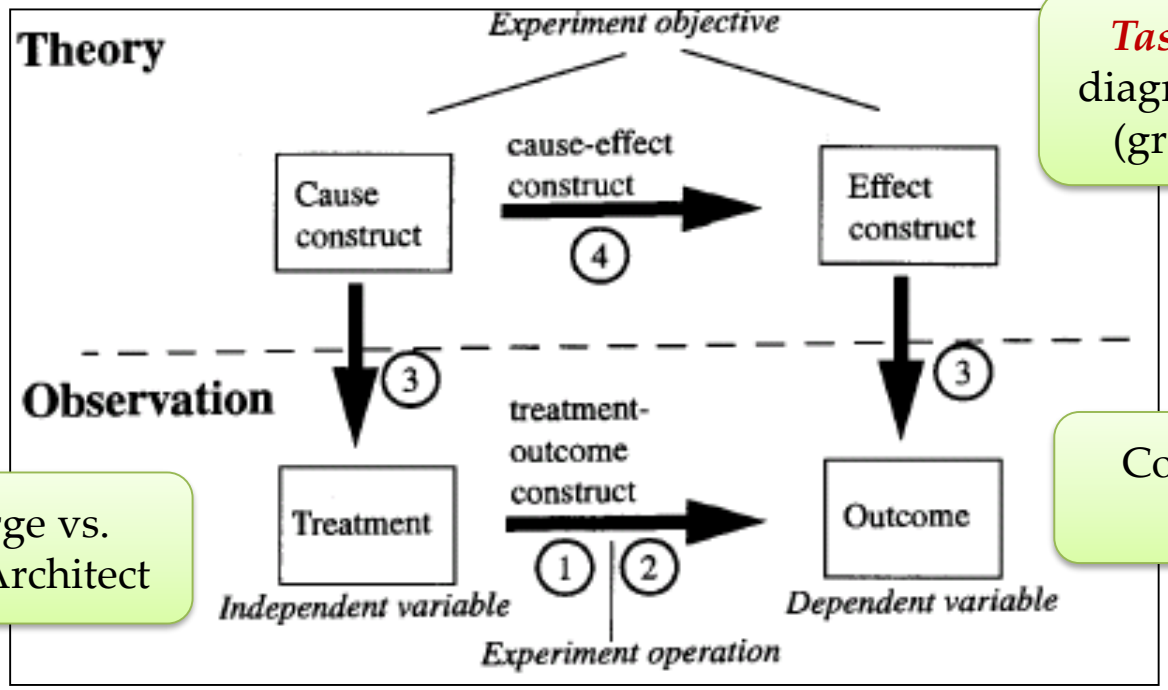
Task (merging class diagrams) and subjects (graduate students)

Stu-merge vs. Rational Architect

Correctness, ***speed***, assessment

Threats to Validity (3)

- ④ external validity (representativeness)
  - Task: class diagram models were of a toy problem
  - Subject: graduate students as sample of what population?



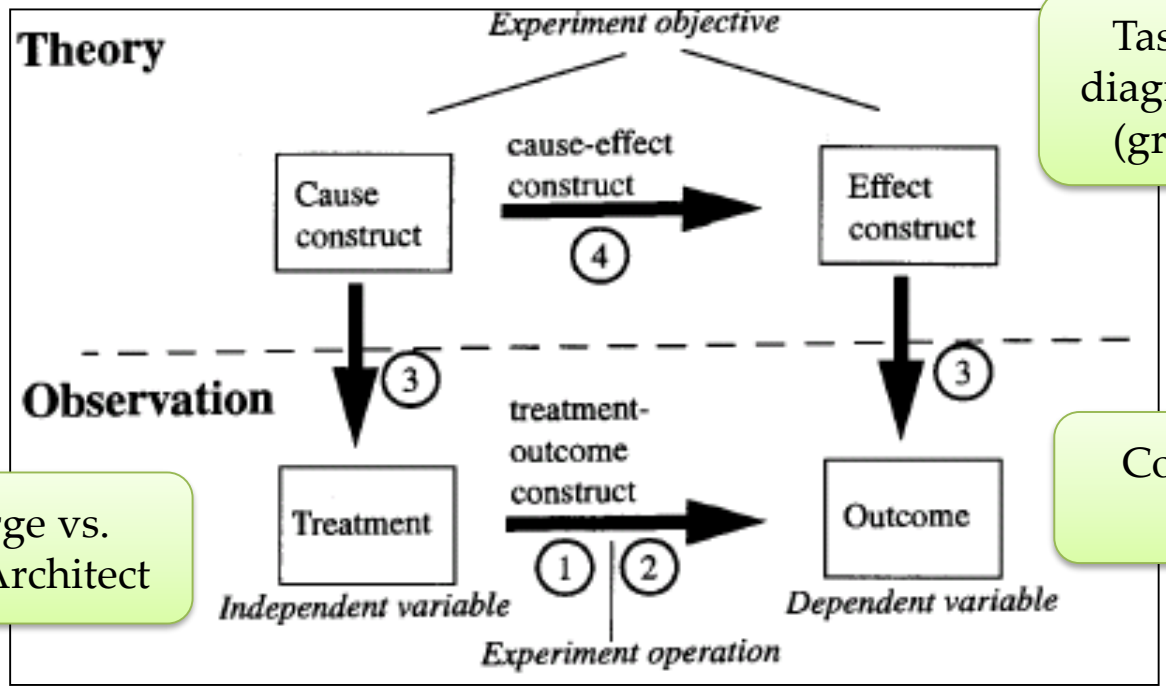
Task (merging class diagrams) and ***subjects*** (graduate students)

Stu-merge vs. Rational Architect

Correctness, speed, assessment

Threats to Validity (4)

- ① conclusion validity (theoretical reliability)
  - Bias: subjects knew Stu-merge was Stu's own tool



Task (merging class diagrams) and subjects (graduate students)

Stu-merge vs. Rational Architect

Correctness, speed, assessment

Exercise

- Describe your experimental design
  - ***Questions*** to answer
  - ***Subjects*** to select
  - ***Independent*** variables, ***dependent*** variables, and ***controlled*** variables (no ***confounding*** variable)

- Answer questions about
  - ***Threats*** to construct validity, internal validity, external validity, and conclusion validity
  - Why do they ***not affect*** your conclusion?