28th International Conference
on Software Engineering

Tutorial F2

# Case Studies for Software Engineers

Steve Easterbrook University of Toronto

Jorge Aranda University of Toronto

This tutorial was originally developed with:
Susan Elliott Sim University of California, Irvine
Dewayne E. Perry The University of Texas at Austin

## Goals of this tutorial

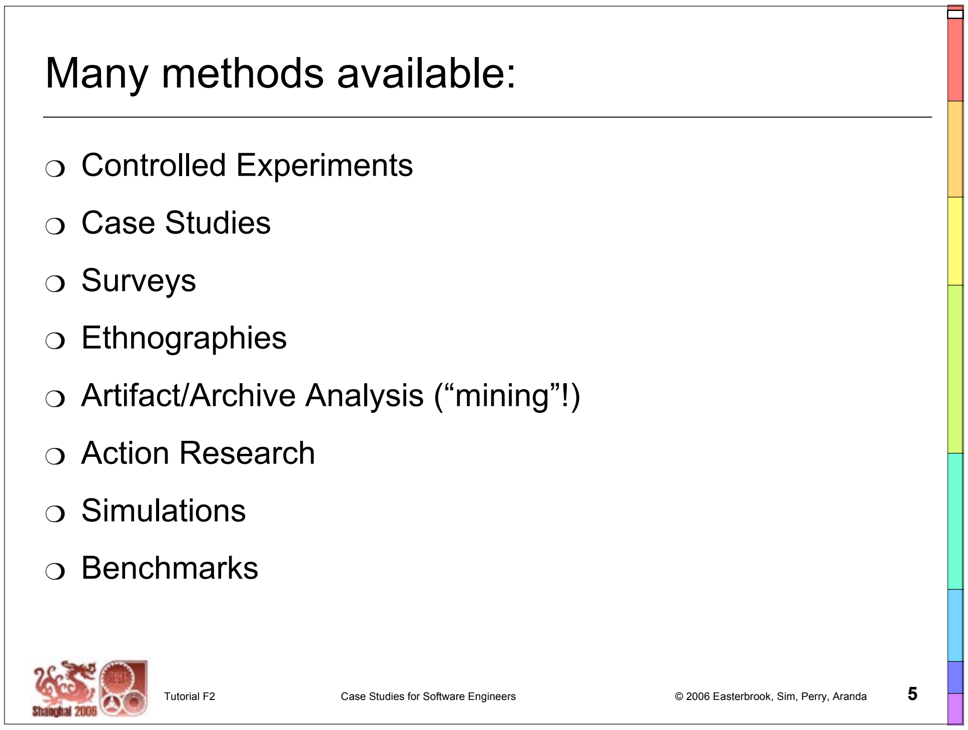
- For researchers:
  - differentiate case studies from other empirical methods
  - a solid grounding in the fundamentals of case studies as a research method
  - understand and avoid common mistakes with case studies
- For reviewers:
  - guidance to judge quality and validity of reported case studies.
  - criteria to assess whether research papers based on case studies are suitable for publication
- For practitioners:
  - awareness of how to interpret the claims made by researchers about new software engineering methods and tools.
  - insights into the roles practitioners can play in case studies research

## Overview

- Session 1:
  Recognising Case Studies
  - ● 9:00-9:30 Empirical Methods in SE
  - ● 9:30-9:45 Basic Elements of a case study
  - 9:45-10:30 *Exercise: Read and Discuss Published Case studies*
  - **10:30-11:00 Coffee break**
- Session 2:
  Designing Case Studies I
  - ● 11:00-11:30 Theory Building
  - 11:30-11:45 *Exercise: Identifying theories in SE*
  - ● 11:45-12:30 Planning and Data Collection
  - **12:30-2:00 Lunch**
- Session 3:
  Designing Case Studies II
  - 2:00-2:30 *Exercise: Design a Case Study*
  - ● 2:30-3:30 Data Analysis and Validity
  - **3:30-4:00 Tea break**
- Session 4:
  Publishing and Reviewing Case Studies
  - 4:00-4:30 *Exercise: Design a Case Study (concludes)*
  - ● 4:30-5:00 Publishing Case Studies
  - ● 5:00-5:15 Reviewing Case Studies
  - ● 5:15-5:30 Summary/Discussion
  - **5:30 Finish**

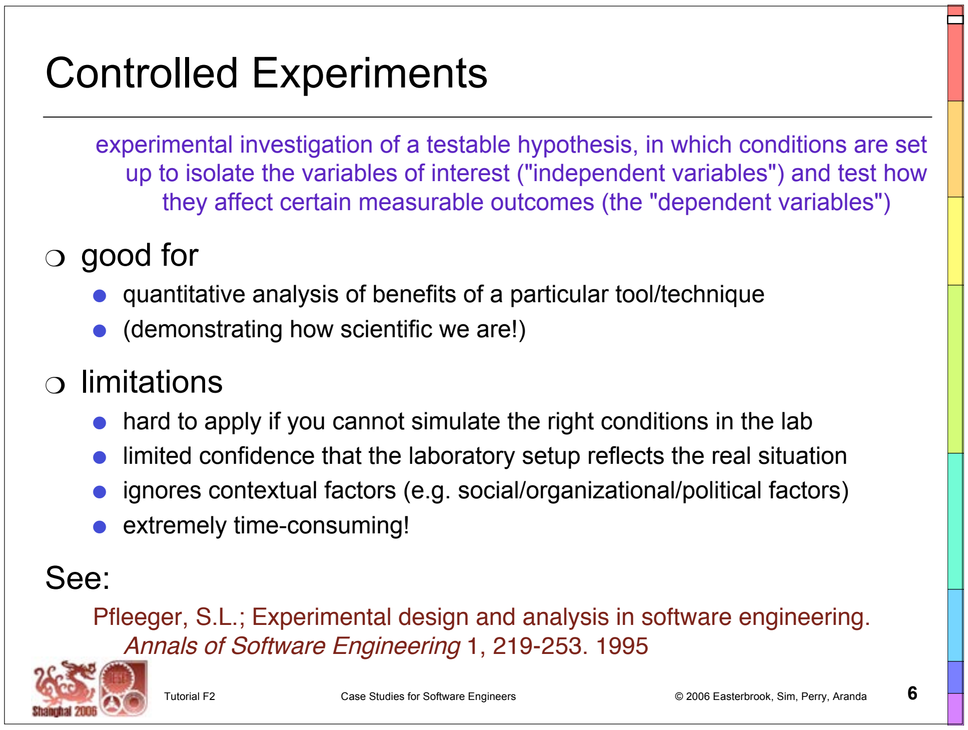
28th International Conference
on Software Engineering

## 1. Empirical Methods in Software Engineering



## Many methods available:

- ❍ Controlled Experiments
- ❍ Case Studies
- ❍ Surveys
- ❍ Ethnographies
- ❍ Artifact/Archive Analysis (“mining”!)
- ❍ Action Research
- ❍ Simulations
- ❍ Benchmarks



## Controlled 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 laboratory setup reflects the real situation
  - ignores contextual factors (e.g. social/organizational/political factors)
  - extremely time-consuming!

See:

Pfleeger, S.L.; Experimental design and analysis in software engineering. *Annals of Software Engineering* 1, 219-253. 1995

## 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

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

## Ethnographies

Interpretive, in-depth studies in which the researcher immerses herself in a social group under study to understand phenomena though 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.

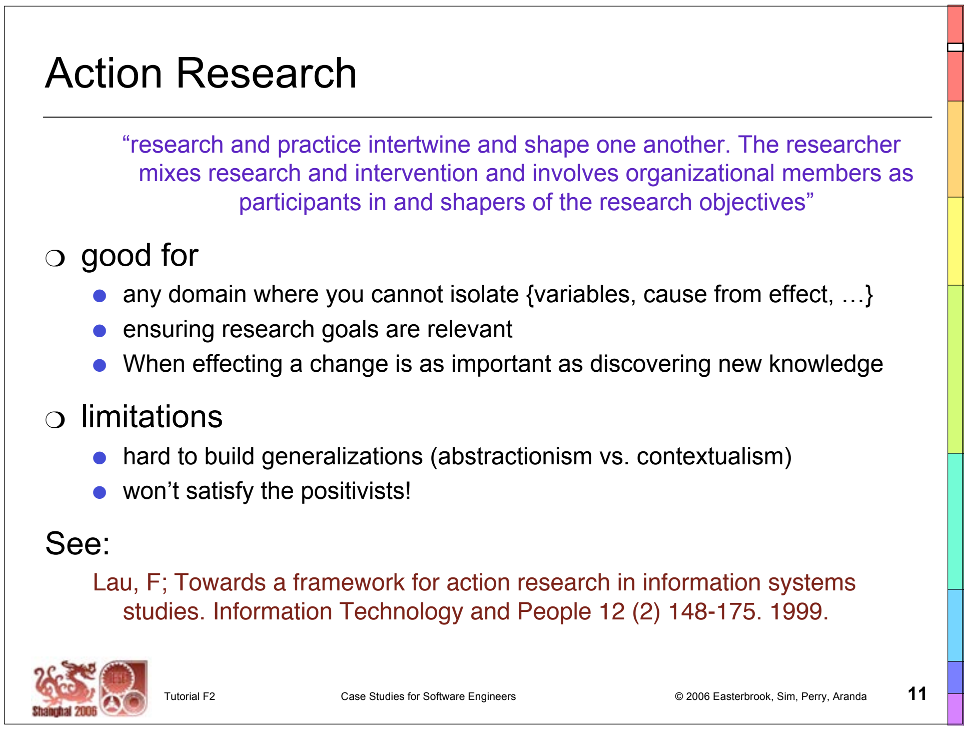
## Artifact / Archive Analysis

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.



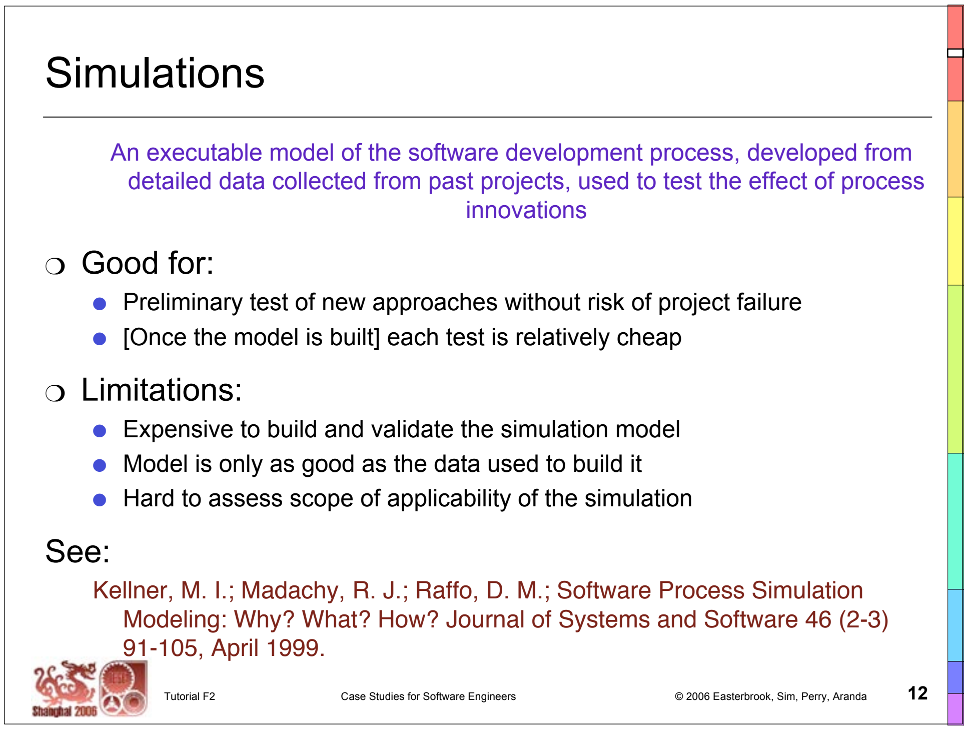
## Action Research

“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. Information Technology and People 12 (2) 148-175. 1999.



## 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

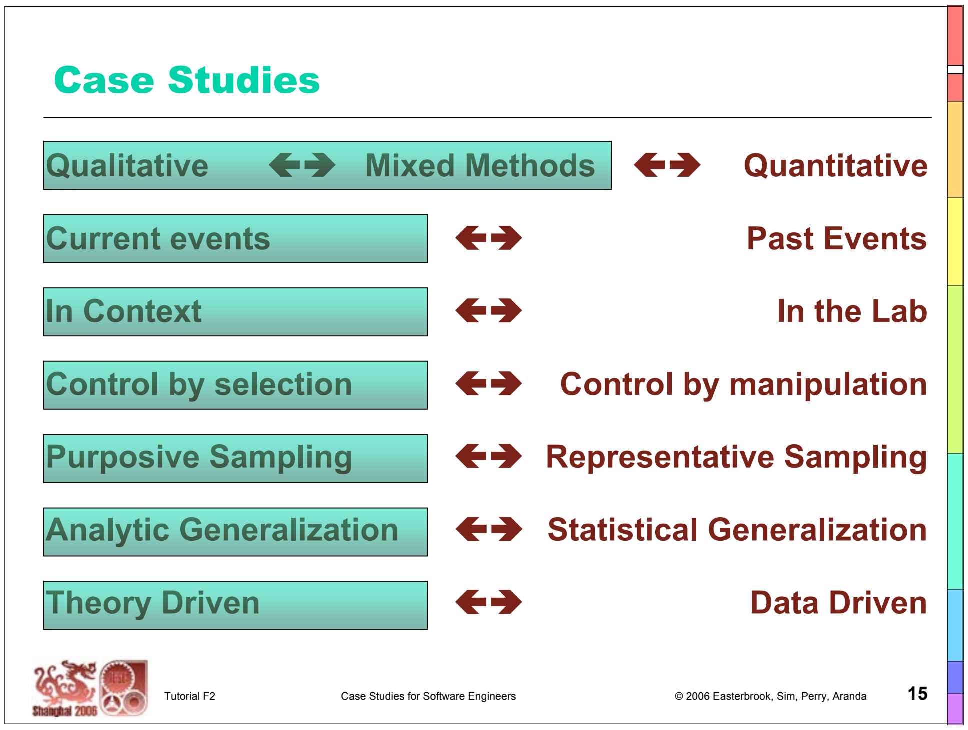
- good for
  - making detailed comparisons between methods/tools
  - increasing the (scientific) maturity of a research community
  - building consensus over the valid problems and approaches to them
- limitations
  - can only be applied if the community is ready
  - become less useful / redundant as 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

## Comparing Empirical Research Methods

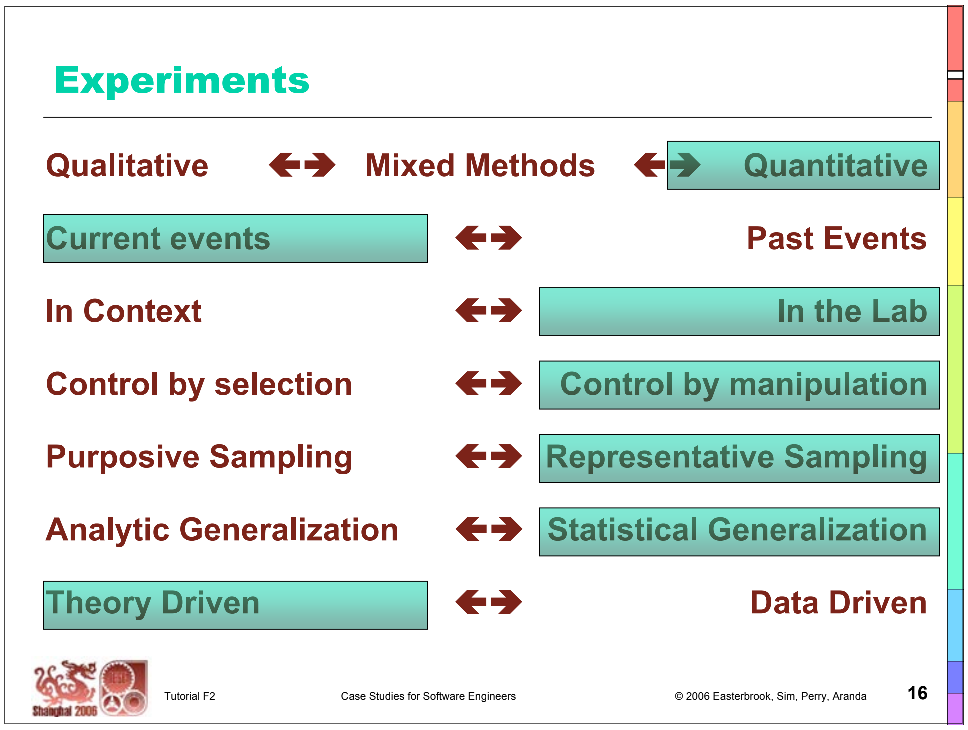
| | | | | |
|---|---|---|---|---|
| **Qualitative** | **⇔** | **Mixed Methods** | **⇔** | **Quantitative** |
| **Current events** | | **⇔** | | **Past Events** |
| **In Context** | | **⇔** | | **In the Lab** |
| **Control by selection** | | **⇔** | | **Control by manipulation** |
| **Purposive Sampling** | | **⇔** | | **Representative Sampling** |
| **Analytic Generalization** | | **⇔** | | **Statistical Generalization** |
| **Theory Driven** | | **⇔** | | **Data Driven** |



## Case Studies

| | | |
|---|---|---|
| Qualitative ⇔ Mixed Methods | ⇔ | Quantitative |
| Current events | ⇔ | Past Events |
| In Context | ⇔ | In the Lab |
| Control by selection | ⇔ | Control by manipulation |
| Purposive Sampling | ⇔ | Representative Sampling |
| Analytic Generalization | ⇔ | Statistical Generalization |
| Theory Driven | ⇔ | Data Driven |

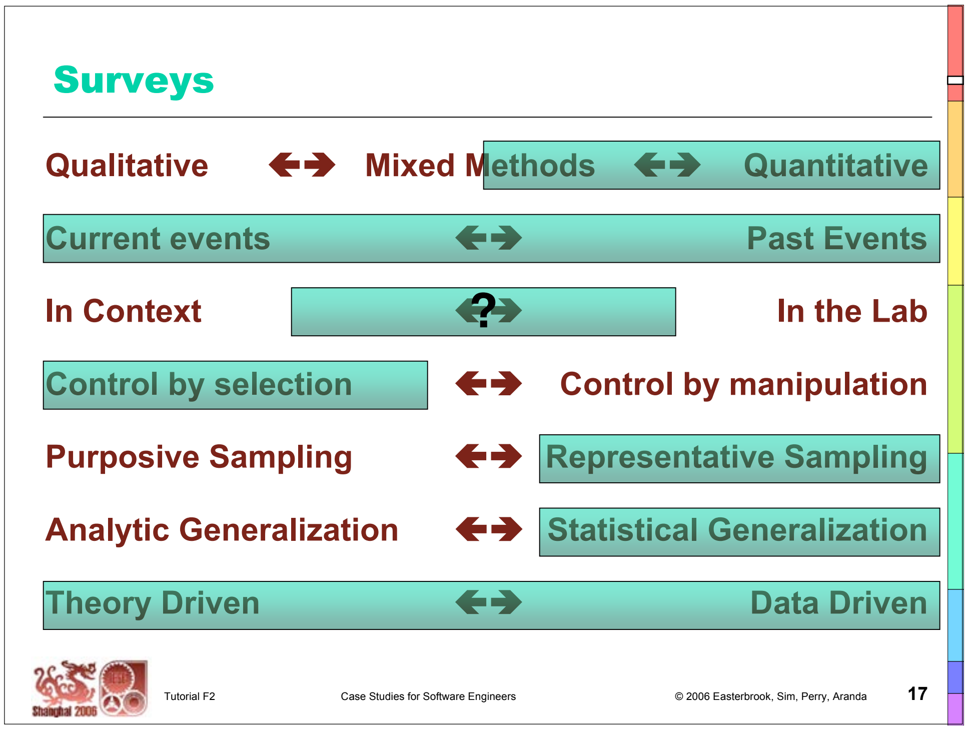
Tutorial F2 Case Studies for Software Engineers © 2006 Easterbrook, Sim, Perry, Aranda 15



## Experiments

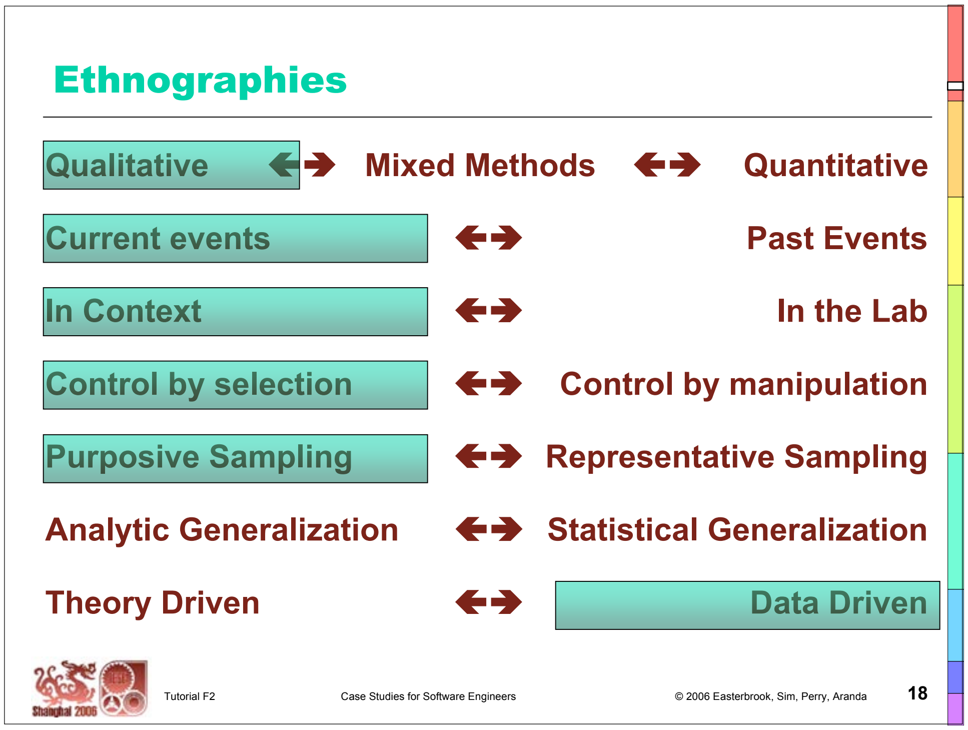
| | | |
|---|---|---|
| Qualitative ⇔ Mixed Methods | ⇔ | Quantitative |
| Current events | ⇔ | Past Events |
| In Context | ⇔ | In the Lab |
| Control by selection | ⇔ | Control by manipulation |
| Purposive Sampling | ⇔ | Representative Sampling |
| Analytic Generalization | ⇔ | Statistical Generalization |
| Theory Driven | ⇔ | Data Driven |

Tutorial F2 Case Studies for Software Engineers © 2006 Easterbrook, Sim, Perry, Aranda 16



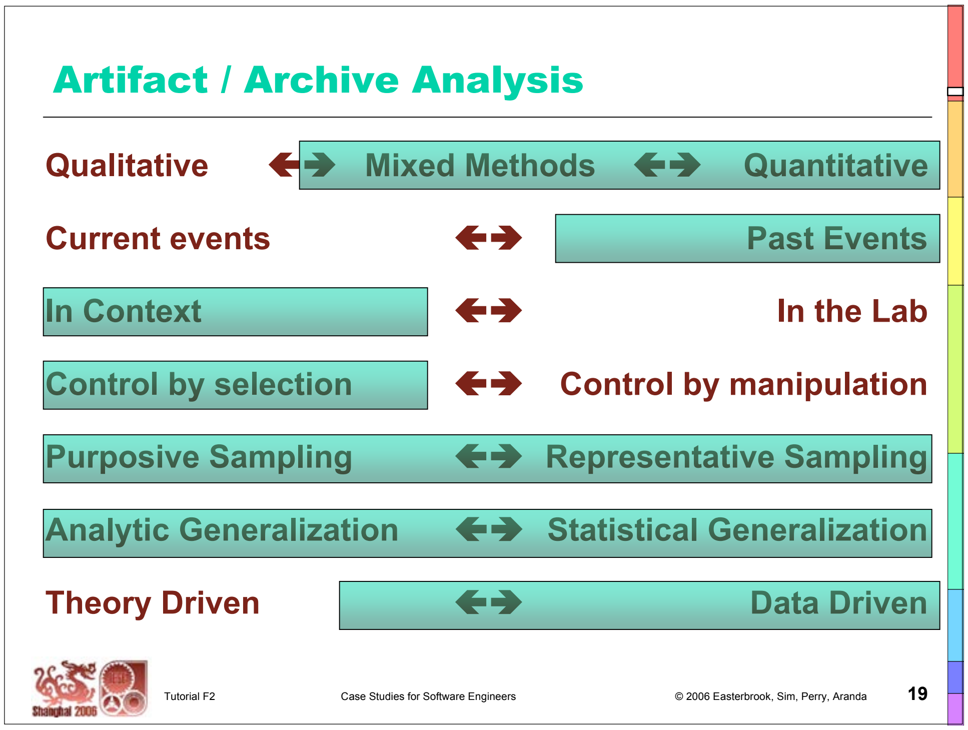
## Surveys

| | | | | |
|---|---|---|---|---|
| Qualitative | ↔ | Mixed Methods | ↔ | Quantitative |
| Current events | ↔ | | | Past Events |
| In Context | ? | | | In the Lab |
| Control by selection | ↔ | | | Control by manipulation |
| Purposive Sampling | ↔ | | | Representative Sampling |
| Analytic Generalization | ↔ | | | Statistical Generalization |
| Theory Driven | ↔ | | | Data Driven |



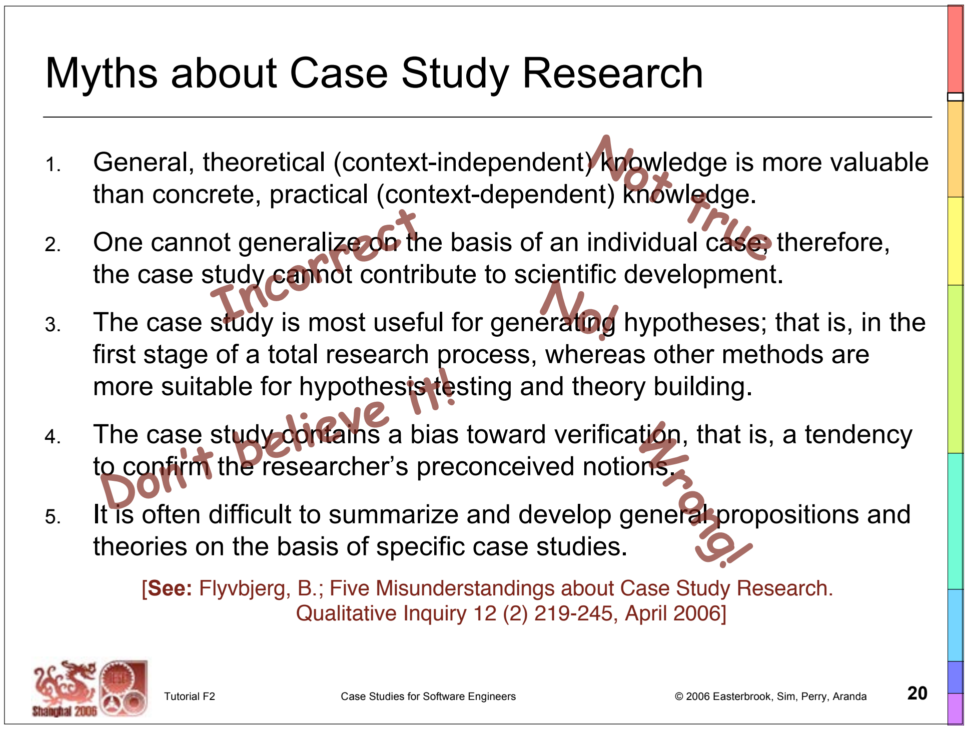
## Ethnographies

| | | | | |
|---|---|---|---|---|
| Qualitative | ↔ | Mixed Methods | ↔ | Quantitative |
| Current events | ↔ | | | Past Events |
| In Context | ↔ | | | In the Lab |
| Control by selection | ↔ | | | Control by manipulation |
| Purposive Sampling | ↔ | | | Representative Sampling |
| Analytic Generalization | ↔ | | | Statistical Generalization |
| Theory Driven | ↔ | | | Data Driven |



## Artifact / Archive Analysis

| | | |
|---|---|---|
| Qualitative | ⇔ Mixed Methods ⇔ | Quantitative |
| Current events | ⇔ | Past Events |
| In Context | ⇔ | In the Lab |
| Control by selection | ⇔ | Control by manipulation |
| Purposive Sampling | ⇔ | Representative Sampling |
| Analytic Generalization | ⇔ | Statistical Generalization |
| Theory Driven | ⇔ | Data Driven |



## Myths about Case Study Research

1. General, theoretical (context-independent) knowledge is more valuable than concrete, practical (context-dependent) knowledge.
2. One cannot generalize on the basis of an individual case; therefore, the case study cannot contribute to scientific development.
3. The case study is most useful for generating hypotheses; that is, in the first stage of a total research process, whereas other methods are more suitable for hypothesis testing and theory building.
4. The case study contains a bias toward verification, that is, a tendency to confirm the researcher's preconceived notions.
5. It is often difficult to summarize and develop general propositions and theories on the basis of specific case studies.

Not true

Incorrect

No!

Don't believe it!

Wrong!

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

28th International Conference
on Software Engineering

## 2. Basic Elements of a Case Study

## Overview

- Differentiating the Case Study Method
- When to use Case Studies
- Types of Case Study
- Key elements of Case Study Design
- Quality Criteria for good Case Study Research

## When should you use a case study?

- When you can't control the variables
- When there are many more variables than data points
- When you cannot separate phenomena from context
  - Phenomena that don't occur in a lab setting
  - E.g. large scale, complex software projects
  - Effects can be wide-ranging.
  - Effects can take a long time to appear (weeks, months, years!)
- When the context is important
  - E.g. When you need to know how context affects the phenomena
- When you need to know whether your theory applies to a specific real world setting

## Why conduct a case study?

- To gain a deep understanding of a phenomenon
  - Example: To understand the capability of a new tool
  - Example: To identify factors affecting communication in code inspections
  - Example: To characterize the process of coming up to speed on a project
- Objective of Investigation
  - Exploration- To find what's out there
  - Characterization- To more fully describe
  - Validation- To find out whether a theory/hypothesis is true
- Subject of Investigation
  - An intervention, e.g. tool, technique, method, approach to design, implementation, or organizational structure
  - An existing thing or process, e.g. a team, releases, defects

## Misuses of the term "Case Study"

- ❍ Not a case history
  - In medicine and law, patients or clients are "cases." Hence sometimes they refer to a review of interesting instance(s) as a "case study".
- ❍ Not an exemplar
  - Not a report of something interesting that was tried on a toy problem
- ❍ Not an experience report
  - Retrospective report on an experience (typically, industrial) with lessons learned
- ❍ Not a quasi-experiment with small n
  - Weaker form of experiment with a small sample size
  - Uses a different logic for designing the study and for generalizing from results

## How can I tell it's a case study?

- ❍ Has research questions set out from the beginning of the study
- ❍ Data is collected in a planned and consistent manner
- ❍ Inferences are made from the data to answer the research questions
- ❍ Produces an explanation, description, or causal analysis of a phenomenon
  - Can also be exploratory
- ❍ Threats to validity are addressed in a systematic way

## Parts of a Case Study Research Design

- A research design is a "blueprint" for a study
  - Deals more with the logic of the study than the logistics
  - Plan for moving from questions to answers
  - Ensures data is collected and analyzed to produce an answer to the initial research question
  - (Analogy: research design is like a system design)
- Five parts of a case study research design
  1. Research questions
  2. Propositions (if any)
  3. Unit(s) of analysis
  4. Logic linking the data to the propositions
  5. Criteria for interpreting the findings

## Part 1: Study Questions

- Study design always starts with research questions
  - Clarify precisely the nature of the research question
  - Ensure the questions can be answered with a case study
  - Generally, should be "how" and "why" questions.
  - Identify and interpret the relevant theoretical constructs
- Examples:
  - "Why do 2 organizations have a collaborative relationship?"
  - "Why do developers prefer this tool/model/notation?"
  - "How are inspections carried out in practice?"
  - "How does agile development work in practice?"
  - "Why do programmers fail to document their code?"
  - "How does software evolve over time?"
  - "Why have formal methods not been adopted widely for safety-critical software?"
  - "How does a company identify which software projects to start?"

## Types of Case Studies

- Explanatory
  - Adjudicates between competing explanations (theories)
  - E.g. How important is implementation bias in requirements engineering?
    - Rival theories: existing architectures are useful for anchoring, vs. existing architectures are over-constraining during RE
- Descriptive
  - Describes sequence of events and underlying mechanisms
  - E.g. How does pair programming actually work?
  - E.g. How do software immigrants naturalize?
- Causal
  - Looks for causal relationship between concepts
  - E.g. How do requirements errors and programming errors affect safety in real time control systems?
    - See study by Robyn Lutz on the Voyager and Galileo spacecraft
- Exploratory
  - Used to build new theories where we don't have any yet
  - Choose cases that meet particular criteria or parameters
  - E.g. Christopher Columbus' voyage to the new world
  - E.g. What do CMM level 3 organizations have in common?

## Part 2: Study Propositions

- Propositions are claims about the research question
  - State what you expect to show in the study
  - Direct attention to things that should be examined in the case study
  - E.g. "Organizations collaborate because they derive mutual benefits"
- Propositions will tell you where to look for relevant evidence
  - Example: Define and ascertain the specific benefits to each organization
- Note: exploratory studies might not have propositions
  - …but should lead to propositions for further study
  - …and should still have a clearly-stated purpose and clearly-stated criteria for success
- Analogy: hypotheses in controlled experiments

## Part 3: Unit of Analysis

- Defines what a "case" is in the case study
  - Choice depends on the primary research questions
  - Choice affects decisions on data collection and analysis
  - Hard to change the unit of analysis once the study has started (but can be done if there are compelling reasons)
  - Note: good idea to use same unit of analysis as previous studies (why?)
- Often many choices:
  - E.g. for an exploratory study of extreme programming:
    - Unit of analysis = individual developer (case study focuses on a person's participation in the project)
    - Unit of analysis = a team (case study focuses on team activities)
    - Unit of analysis = a decision (case study focuses on activities around that decision)
    - Unit of analysis = a process (e.g. case study examines how user stories are collected and prioritized)
    - …

## Examples of Units of Analysis

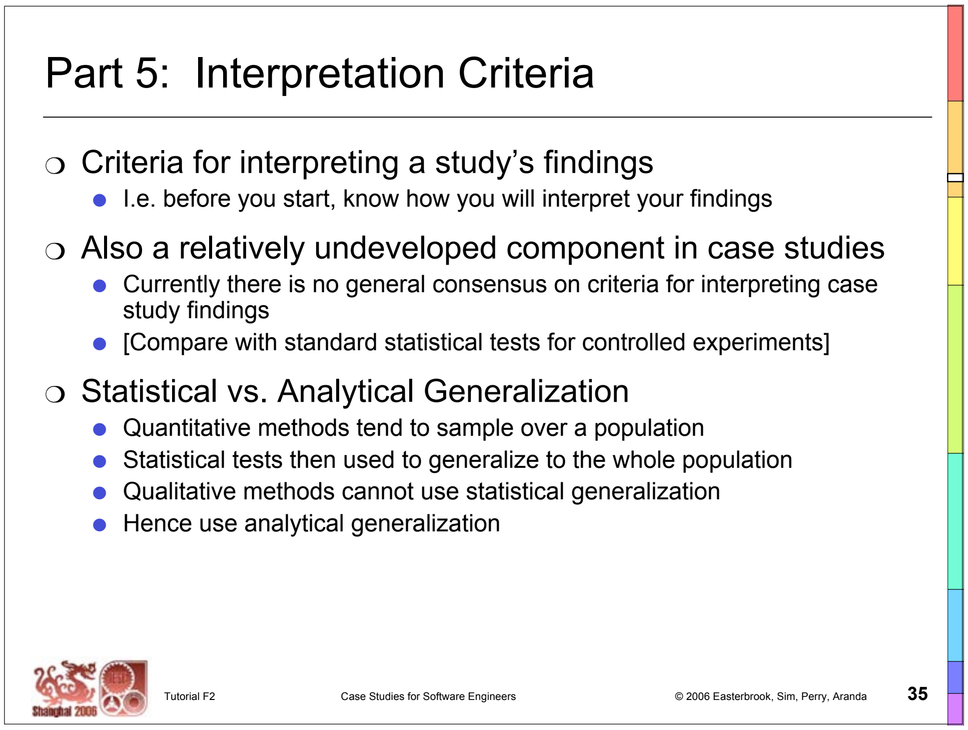
- For a study of how software immigrants naturalize
  - Individuals?
  - … or the Development team?
  - … or the Organization?
- For a study of pair programming
  - Programming episodes?
  - … or Pairs of programmers?
  - … or the Development team?
  - … or the Organization?
- For a study of software evolution
  - A Modification report?
  - … or a File?
  - … or a System?
  - … or a Release?
  - … or a Stable release?

## Why Defining your Unit of Analysis matters

- ❍ Clearly bounds the case study
  - …and tells you which data to collect
- ❍ Makes it easier to compare case studies
  - …incomparable unless you know the units of analysis are the same
- ❍ Avoid subjective judgment of scope:
  - e.g. disagreement about the beginning and end points of a process
- ❍ Avoids mistakes in inferences from the data
  - E.g. If your study proposition talks about team homogeneity…
  - …Won't be able to say much if your units of analysis are individuals

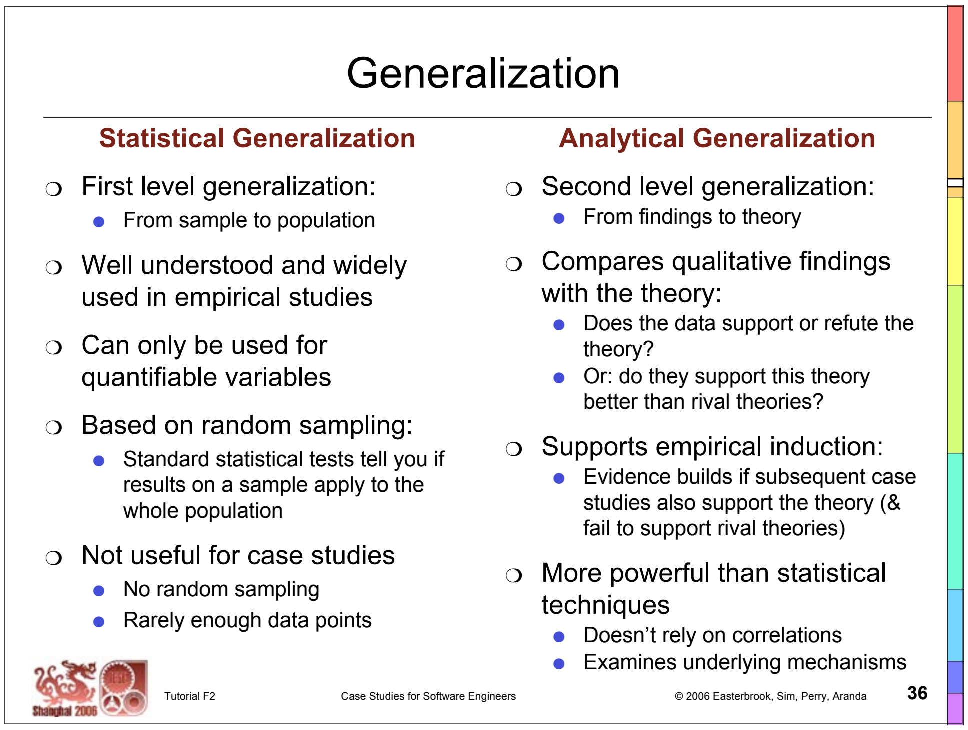
## Part 4: Linking Logic

- ❍ Logic or reasoning to link data to propositions
- ❍ One of the least well developed components in case studies
- ❍ Many ways to perform this
  - …none as precisely defined as the treatment/subject approach used in controlled experiments
- ❍ One possibility is pattern matching
  - Describe several potential patterns, then compare the case study data to the patterns and see which one is closer



## Part 5: Interpretation Criteria

- ❍ Criteria for interpreting a study's findings
  - I.e. before you start, know how you will interpret your findings
- ❍ Also a relatively undeveloped component in case studies
  - Currently there is no general consensus on criteria for interpreting case study findings
  - [Compare with standard statistical tests for controlled experiments]
- ❍ Statistical vs. Analytical Generalization
  - Quantitative methods tend to sample over a population
  - Statistical tests then used to generalize to the whole population
  - Qualitative methods cannot use statistical generalization
  - Hence use analytical generalization



## Generalization

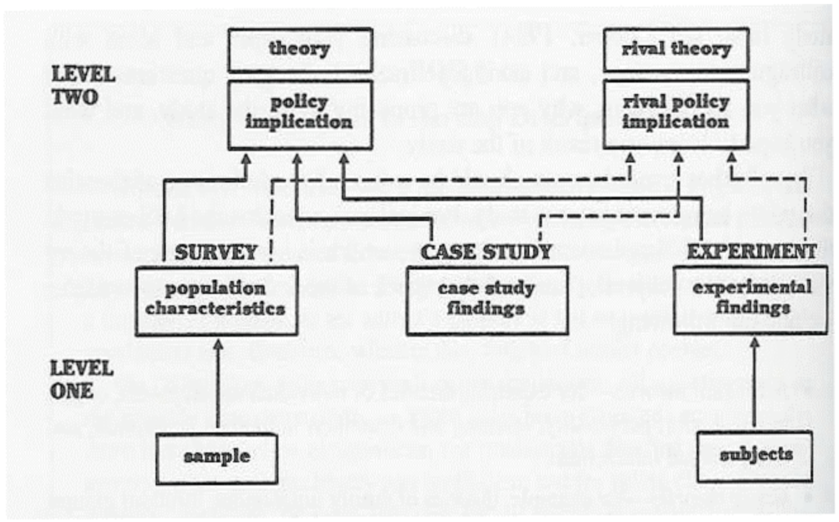
**Statistical Generalization**

- ❍ First level generalization:
  - From sample to population
- ❍ Well understood and widely used in empirical studies
- ❍ Can only be used for quantifiable variables
- ❍ Based on random sampling:
  - Standard statistical tests tell you if results on a sample apply to the whole population
- ❍ Not useful for case studies
  - No random sampling
  - Rarely enough data points

**Analytical Generalization**

- ❍ Second level generalization:
  - From findings to theory
- ❍ Compares qualitative findings with the theory:
  - Does the data support or refute the theory?
  - Or: do they support this theory better than rival theories?
- ❍ Supports empirical induction:
  - Evidence builds if subsequent case studies also support the theory (& fail to support rival theories)
- ❍ More powerful than statistical techniques
  - Doesn't rely on correlations
  - Examines underlying mechanisms

## Analytical and Statistical Generalization



## How can I evaluate a case study?

Same criteria as for other empirical research:

- Construct Validity
  - Concepts being studied are operationalized and measured correctly
- Internal Validity
  - Establish a causal relationship and distinguish spurious relationships
- External Validity
  - Establish the domain to which a study's findings can be generalized
- Empirical Reliability
  - Demonstrate that the study can be repeated with the same results

28th International Conference
on Software Engineering

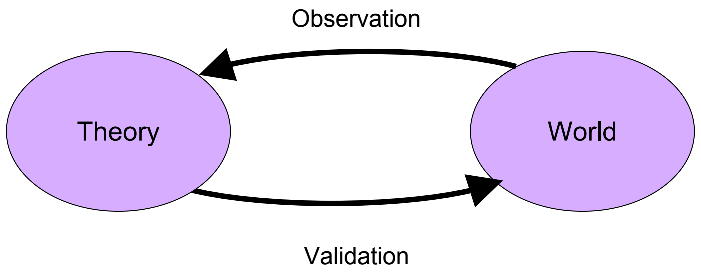
## Exercise:
## Read and Discuss Published Case Studies

28th International Conference
on Software Engineering

## Theory Building

## Scientific Method

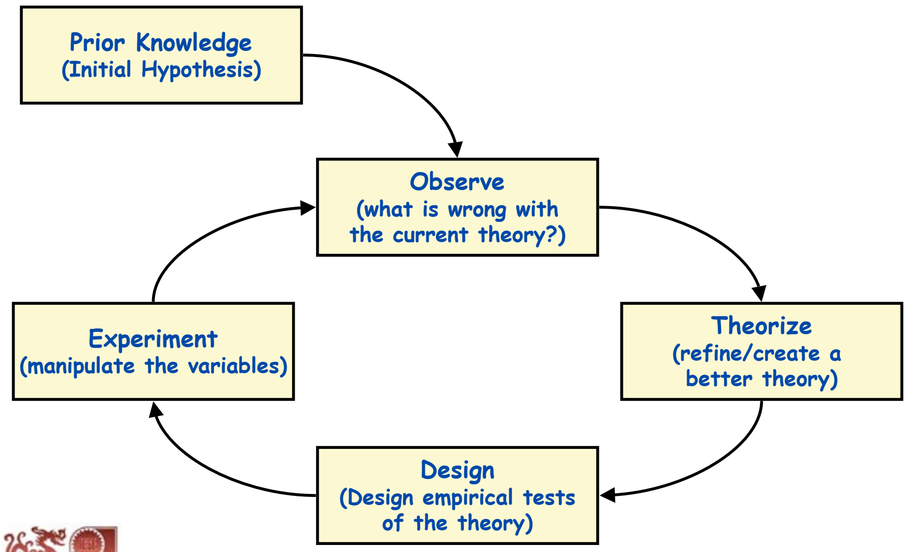
- ❍ No single "official" scientific method
  http://dharma-haven.org/science/myth-of-scientific-method.htm
- ❍ However, there are commonalities



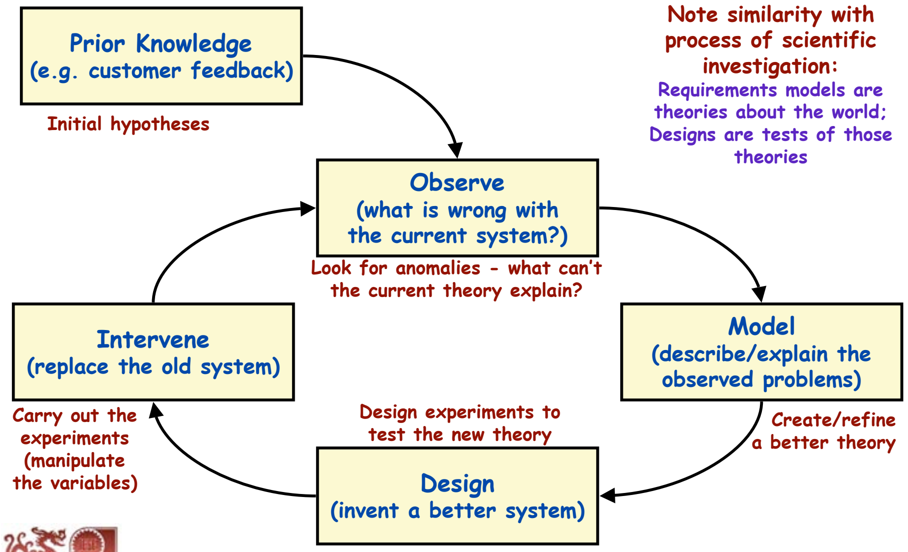
## High School Science Version

1. Observe some aspect of the universe.
2. Invent a tentative description, called a *hypothesis*, that is consistent with what you have observed.
3. Use the hypothesis to make predictions.
4. Test those predictions by experiments or further observations and modify the hypothesis in the light of your results.
5. Repeat steps 3 and 4 until there are no discrepancies between theory and experiment and/or observation.

## Scientific Inquiry


Prior Knowledge
(Initial Hypothesis)
Observe
(what is wrong with the current theory?)
Theorize
(refine/create a better theory)
Design
(Design empirical tests of the theory)
Experiment
(manipulate the variables)

## (Comparison: The Engineering Cycle)


Prior Knowledge
(e.g. customer feedback)
Initial hypotheses
Observe
(what is wrong with the current system?)
Look for anomalies - what can't the current theory explain?
Model
(describe/explain the observed problems)
Create/refine a better theory
Design
(invent a better system)
Design experiments to test the new theory
Intervene
(replace the old system)
Carry out the experiments (manipulate the variables)

Note similarity with process of scientific investigation:

Requirements models are theories about the world; Designs are tests of those theories

## 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 as important as in art
- Theories, hypotheses, experimental designs
- Search for elegance, simplicity

## 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 provides an explanation of a set of phenomena
- Ideally, the theory has predictive power too

❍ A hypothesis is a testable statement that is derived from a theory
- A hypothesis is not a theory!

In software engineering, there are few "Theories"
- Many "small-t" theories, philosophers call these *folk theories*

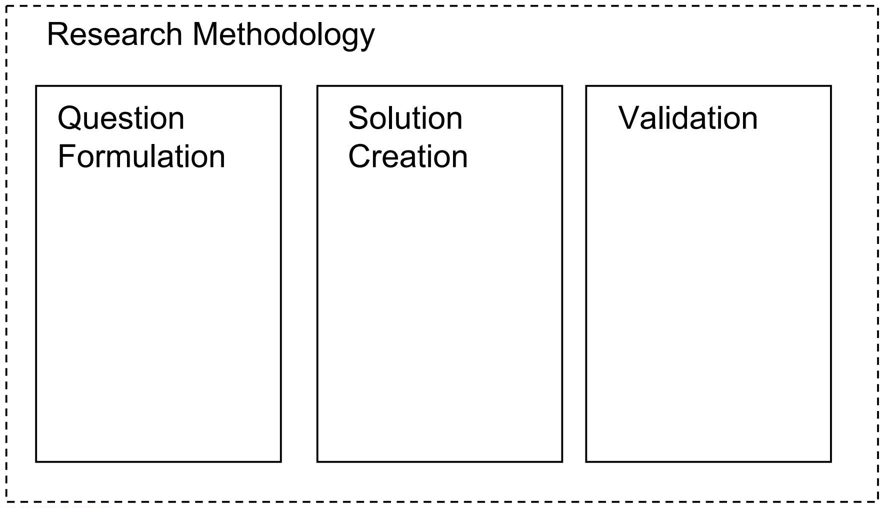
## Science and Theory

- A (scientific) Theory is:
  - more than just a description - it explains and predicts
  - Logically complete, internally consistent, falsifiable
  - Simple and elegant.
- Components
  - concepts, relationships, causal inferences
    - E.g. Conway's Law- structure of software reflects the structure of the team that builds it. A theory should explain why.
- Theories lie at the heart of what it means to do science.
  - Production of generalizable knowledge
  - Scientific method <-> Research Methodology <-> Proper Contributions for a Discipline
- Theory provides orientation for data collection
  - Cannot observe the world without a theoretical perspective

## Meta-theories (theories about theory)

- 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
- Lakatos:
  - Not one paradigm, but many competing research programmes
  - 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 ok
- 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

## Empirical Approach



## Empirical Approaches

- Three approaches
  - Descriptive
  - Relational
  - Experimental/Analytical

## Empirical Approaches

❍ Descriptive

- Goal: careful mapping out a situation in order to describe what is happening
- Necessary first step in any research
  - Provides the basis or cornerstone
  - Provides the what
- Rarely sufficient – often want to know why or how
- But often provides the broad working hypothesis

## Empirical Approaches

❍ Relational

- Need at least two sets of observations so that some phenomenon can be related to each other
- Two or more variables are measured and related to each other
- Coordinated observations -> quantitative degree of correlation
- Not sufficient to explain why there is a correlation

## Empirical Approaches

- ❍ Experimental/Analytical
  - Focus on identification of cause and effect — what leads to what?
  - Want "**X** is responsible for **Y"**, not "**X** is related to **Y"**
  - Need to manipulate the variables
- ❍ Two approaches:
  - Set up treatments with different values of key variables
  - Select existing cases with "theoretically" interesting properties
- ❍ Many potential problems
  - Cause-and-effect is hard to prove conclusively

## Concepts and Terminology

- ❍ Aspects of empirical reasoning
  - Empirical principles: accepted truths justified on the basis of observations
  - Deductive-statistical reasoning – universal laws
  - Inductive-statistical reasoning – probabilistic assertions
    - They deal with uncertainty
    - They are not absolute, invariant rules of nature
- ❍ Behavioral sciences are not sufficient to determine exactitude
  - Human values and individual states of mind
  - Unique nature of the situation which is usually not static
  - Historical and social factors

## Validity

- In software engineering, we worry about various issues:
  - Validation – is the software doing what is needed?
  - is it doing it in an acceptable or appropriate way?
  - Verification – is it doing what the specification stated?
  - are the structures consistent with the way it should perform?
- In empirical work, worried about similar kinds of things
  - Are we testing what we mean to test
  - Are the results due solely to our manipulations
  - Are our conclusions justified
  - What are the results applicable to
- The questions correspond to different *validity* concerns
  - The logic of demonstrating causal connections
  - The logic of evidence

28th International Conference on Software Engineering

## Exercise: Identifying theories in SE

28th International Conference
on Software Engineering

## Planning and Data Collection

## Elements of a Case Study

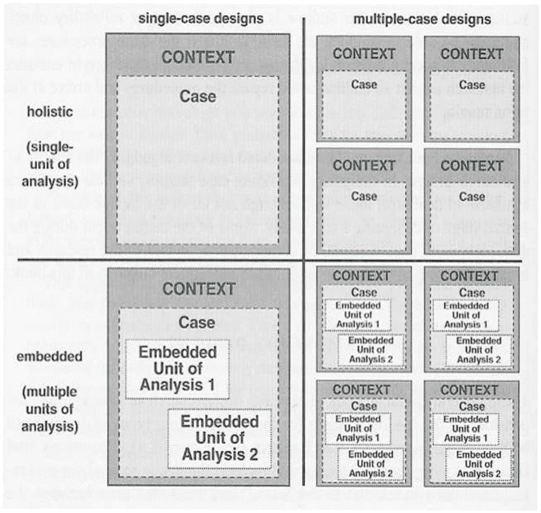
- Research questions
- Propositions (if any)
- Unit(s) of analysis
- Logic linking the data to the propositions
- Criteria for interpreting the findings

## Types of Research Question

- Existence:
  - Does X exist?
- Description and Classification
  - What is X like?
  - What are its properties?
  - How can it be categorized?
- Composition
  - What are the components of X?
- Relationship
  - Are X and Y (cor)related?
- Descriptive-Comparative
  - How does X differ from Y?
- Causality
  - 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?
- Causality-Comparative Interaction
  - Does X cause more Y than does Z under one condition but not others?
- Notes:
  - Don't confuse "causes" & "enables"
  - Don't confuse "causes" & "correlates"

## Case Study Designs

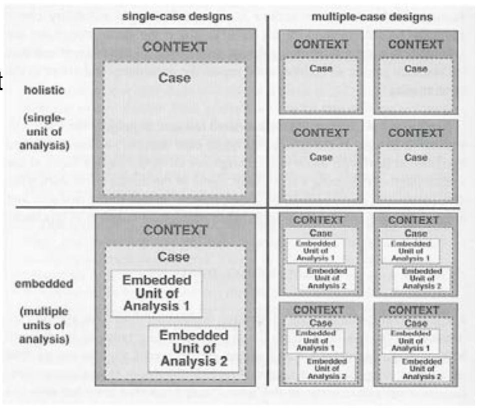
- 4 types of designs (based on a 2x2 matrix)
  - Single-case vs. Multiple-case design
  - Holistic vs. Embedded design


single-case designs
multiple-case designs
holistic
(single-unit of analysis)
embedded
(multiple units of analysis)
CONTEXT
Case
Embedded Unit of Analysis 1
Embedded Unit of Analysis 2

**Basic Types of Designs for Case Studies (Yin, page 40)**

## Holistic vs. Embedded Case Studies

- *Holistic* case study: Examines only the global nature of one unit of analysis (not any subunits)
  - E.g: a case study about an organization
- *Embedded* case study: Involves more than one unit of analysis by paying attention to subunit(s) within the case
  - E.g: a case study about a single organization may have conclusions about the people (subunits) within the organization



single-case designs
multiple-case designs
holistic
(single-unit of analysis)
embedded
(multiple units of analysis)
CONTEXT
Case
Embedded Unit of Analysis 1
Embedded Unit of Analysis 2

## Holistic Designs

- Strengths
  - Convenient when no logical subunits can be defined
  - Good when the relevant theory underlying the case study is holistic in nature
- Weaknesses
  - Can lead to abstract studies with no clear measures or data
  - Harder to detect when the case study is shifting focus away from initial research questions

## Embedded Designs

❍ Strengths
- Introduces higher sensitivity to "slippage" from the original research questions

❍ Weaknesses
- Can lead to focusing only on the subunit (i.e. a multiple-case study of the subunits) and failure to return to the larger unit of analysis

## Rationale for Single-Case Designs

❍ As you might guess, a single-case design uses a single case study to address the research questions

❍ 5 reasons to use a single-case design
- It represents the *critical case* in testing a well-formulated theory
  - The case meets all of the conditions for testing the theory thoroughly
- It represents an *extreme* or *unique* case
  - Example: a case with a rare disorder
- It is the *representative* or *typical* case, i.e. informs about common situations/experiences
  - Gain insights on commonplace situations
- The case is *revelatory* –a unique opportunity to study something previously inaccessible to observation
  - Opens a new topic for exploration
- The case is *longitudinal* – it studies the same case at several points in time
  - The corresponding theory should deal with the change of conditions over time

## Multiple-Case Designs

- Useful when literal or theoretical replications provide valuable information
- Advantages
  - Evidence is considered more compelling
  - Overall study is therefore regarded as more robust
- Disadvantages
  - Difficulty to find an appropriate number of relevant cases
  - Can require extensive resources and time

## Replication in Multiple-Case Studies

- Select each case so that it either:
  - Predicts similar results (*literal replication*)
  - Predicts contrasting results but for predictable reasons (*theoretical replication*)
- If all cases turn out as predicted, there is compelling support for the initial propositions
  - Otherwise the propositions must be revised and retested with another set of cases
- The theoretical framework of the study should guide the choices of replication cases

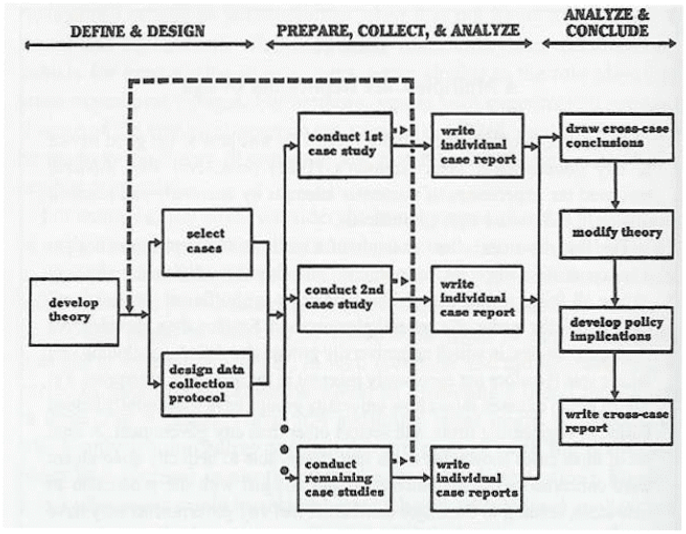
## How Many Cases?

- ❍ How many literal replications?
  - It depends on the certainty you want to have about your results
  - Greater certainty with a larger number of cases
    - Just as with statistical significance measures
    - 2 or 3 may be sufficient if they address very different rival theories and the degree of certainty required is not high
    - 5, 6, or more may be needed for higher degree of certainty
- ❍ How many theoretical replications?
  - Consider the complexity of the domain under study
    - If you are uncertain whether external conditions will produce different results, you may want to include more cases that cover those conditions
    - Otherwise, a smaller number of theoretical replications may be used

## Replication Logic vs. Sampling Logic

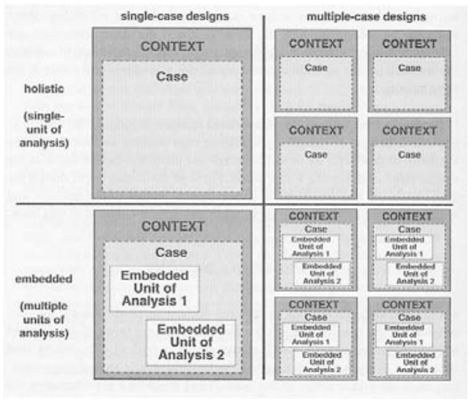
- ❍ Consider multiple-cases analogous to multiple experiments
  - Not analogous to multiple subjects in a single experiment!
- ❍ *Replication logic* (used in case studies) is different from *sampling logic* (used in surveys)
  - Sampling logic requires defining a pool of potential respondents, then selecting a subset using a statistical procedure
  - Responses from the subset are supposed to accurately reflect the responses of the entire pool
- ❍ Sampling logic does not fit with case studies
  - Case studies are not the best method for assessing the prevalence of phenomenon in a population
  - Case studies would have to cover both the phenomenon of interest *and* its context
    - Too many variables, which leads to way too many cases!

## Replication Approach for Multiple-Case Studies

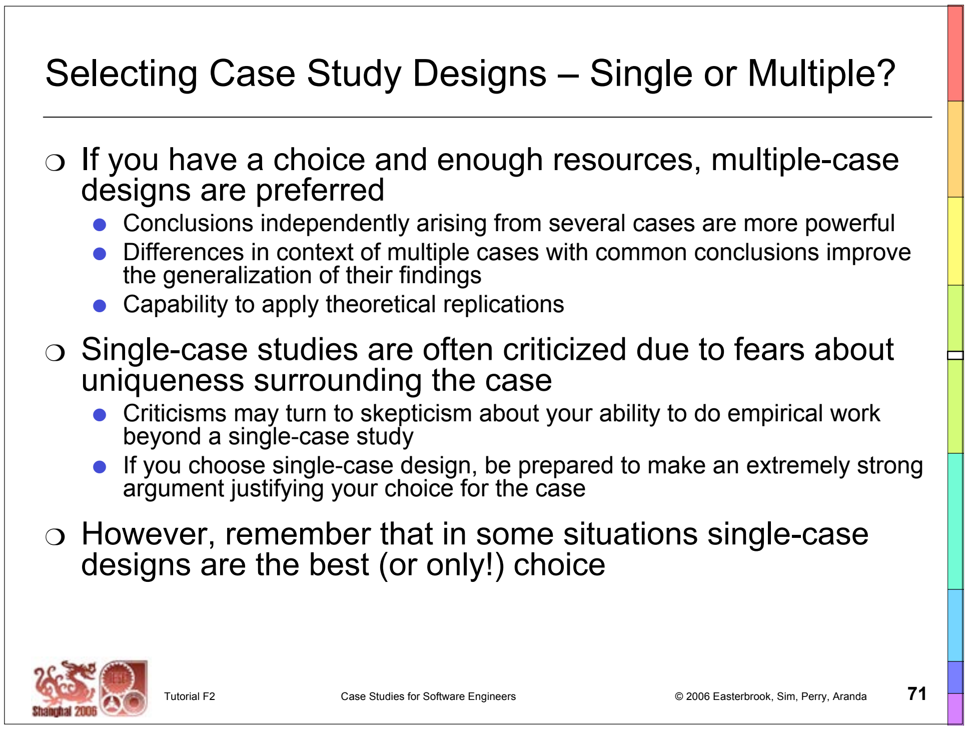

DEFINE & DESIGN
PREPARE, COLLECT, & ANALYZE
ANALYZE & CONCLUDE
develop theory
select cases
design data collection protocol
conduct 1st case study
conduct 2nd case study
conduct remaining case studies
write individual case report
write individual case report
write individual case reports
draw cross-case conclusions
modify theory
develop policy implications
write cross-case report

**Case Study Method (Yin page 50)**

## Multiple-Case Designs: Holistic or Embedded

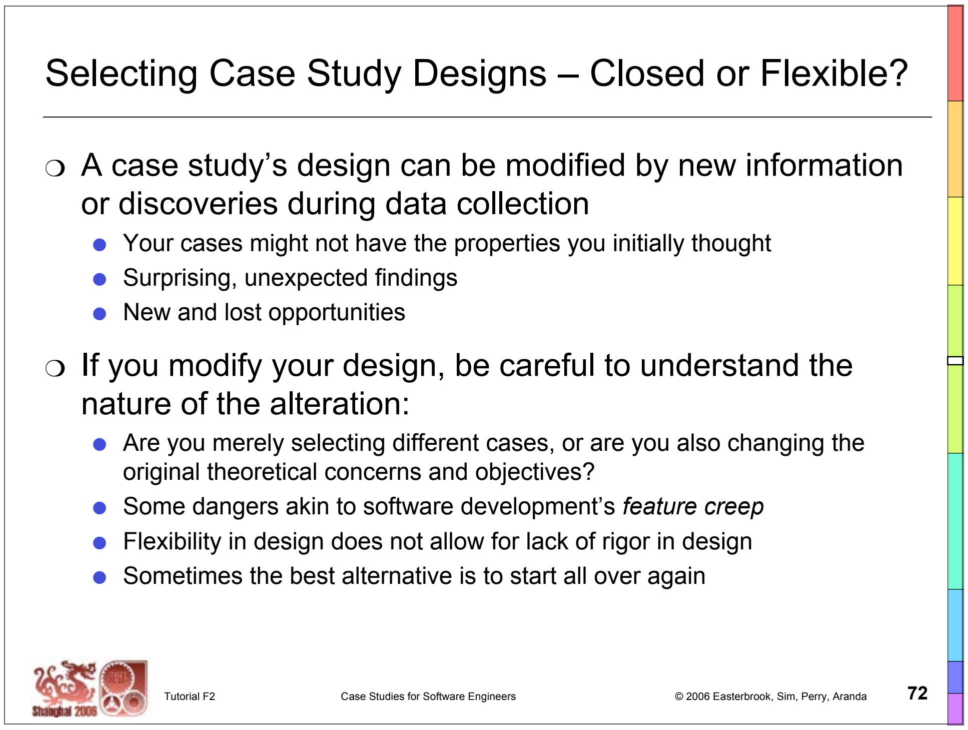

single-case designs
multiple-case designs
holistic
(single-unit of analysis)
embedded
(multiple units of analysis)
CONTEXT
Case
Embedded Unit of Analysis 1
Embedded Unit of Analysis 2

- A multiple-case study can consist of multiple holistic or multiple embedded cases
  - But there is no mixing of embedded and holistic cases in the same study
- Note that for embedded studies, subunit data are *not* pooled across subunits
  - Used to draw conclusions only for the subunit's case



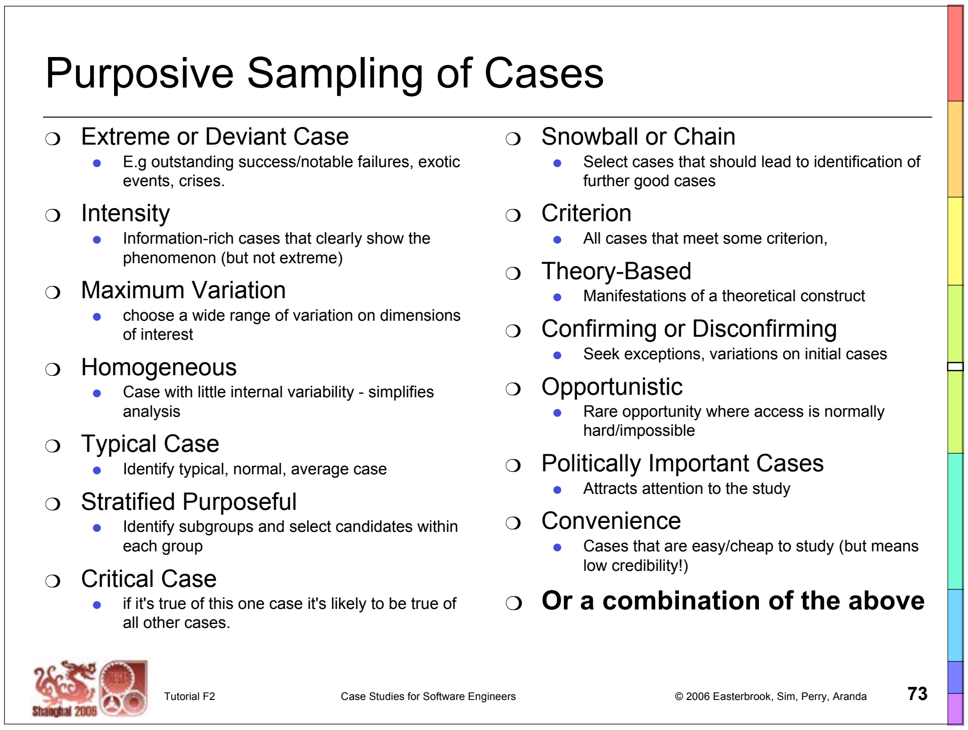
## Selecting Case Study Designs – Single or Multiple?

- ❍ If you have a choice and enough resources, multiple-case designs are preferred
  - Conclusions independently arising from several cases are more powerful
  - Differences in context of multiple cases with common conclusions improve the generalization of their findings
  - Capability to apply theoretical replications
- ❍ Single-case studies are often criticized due to fears about uniqueness surrounding the case
  - Criticisms may turn to skepticism about your ability to do empirical work beyond a single-case study
  - If you choose single-case design, be prepared to make an extremely strong argument justifying your choice for the case
- ❍ However, remember that in some situations single-case designs are the best (or only!) choice



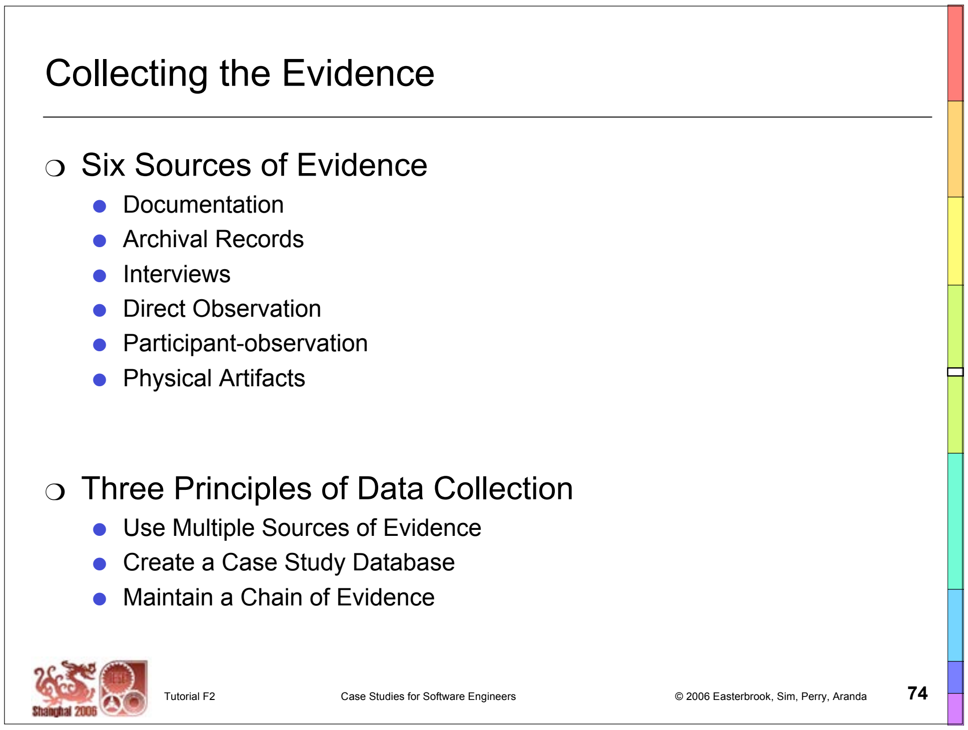
## Selecting Case Study Designs – Closed or Flexible?

- ❍ A case study’s design can be modified by new information or discoveries during data collection
  - Your cases might not have the properties you initially thought
  - Surprising, unexpected findings
  - New and lost opportunities
- ❍ If you modify your design, be careful to understand the nature of the alteration:
  - Are you merely selecting different cases, or are you also changing the original theoretical concerns and objectives?
  - Some dangers akin to software development’s *feature creep*
  - Flexibility in design does not allow for lack of rigor in design
  - Sometimes the best alternative is to start all over again



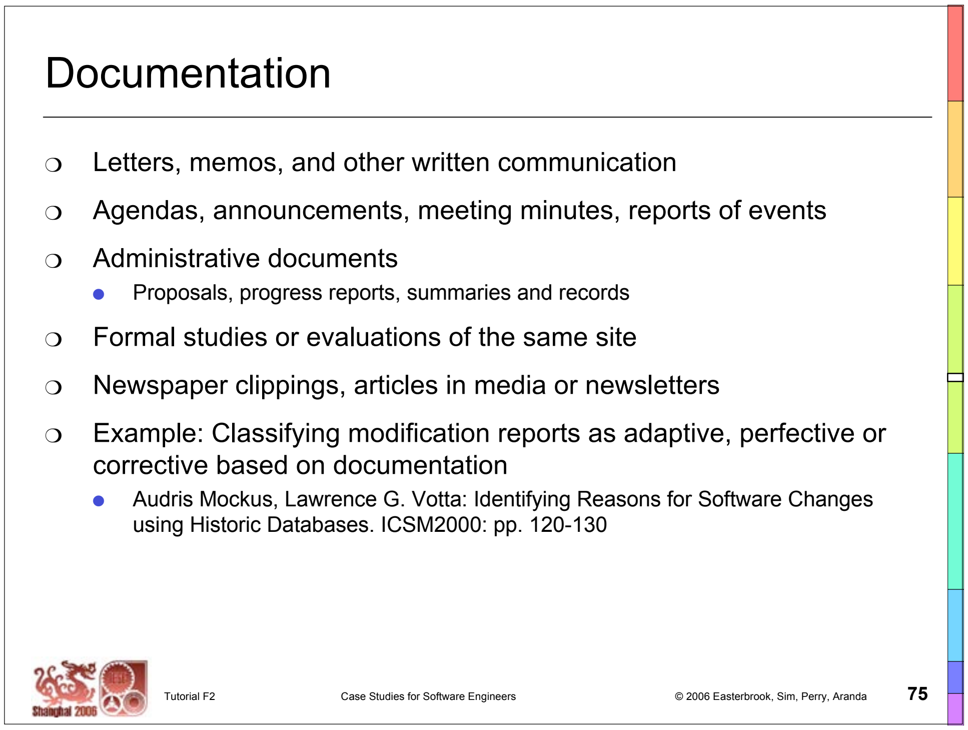
## Purposive Sampling of Cases

- Extreme or Deviant Case
  - E.g outstanding success/notable failures, exotic events, crises.
- Intensity
  - Information-rich cases that clearly show the phenomenon (but not extreme)
- Maximum Variation
  - choose a wide range of variation on dimensions of interest
- Homogeneous
  - Case with little internal variability - simplifies analysis
- Typical Case
  - Identify typical, normal, average case
- Stratified Purposeful
  - Identify subgroups and select candidates within each group
- Critical Case
  - if it's true of this one case it's likely to be true of all other cases.
- Snowball or Chain
  - Select cases that should lead to identification of further good cases
- Criterion
  - All cases that meet some criterion,
- Theory-Based
  - Manifestations of a theoretical construct
- Confirming or Disconfirming
  - Seek exceptions, variations on initial cases
- Opportunistic
  - Rare opportunity where access is normally hard/impossible
- Politically Important Cases
  - Attracts attention to the study
- Convenience
  - Cases that are easy/cheap to study (but means low credibility!)
- **Or a combination of the above**



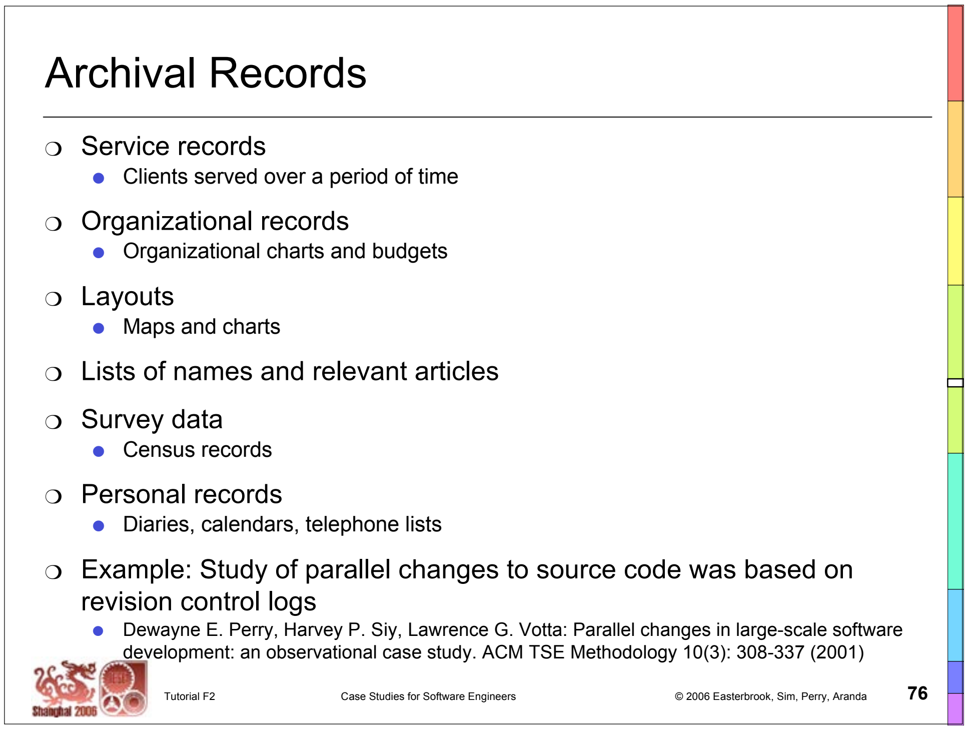
## Collecting the Evidence

- Six Sources of Evidence
  - Documentation
  - Archival Records
  - Interviews
  - Direct Observation
  - Participant-observation
  - Physical Artifacts
- Three Principles of Data Collection
  - Use Multiple Sources of Evidence
  - Create a Case Study Database
  - Maintain a Chain of Evidence



## Documentation

- Letters, memos, and other written communication
- Agendas, announcements, meeting minutes, reports of events
- Administrative documents
  - Proposals, progress reports, summaries and records
- Formal studies or evaluations of the same site
- Newspaper clippings, articles in media or newsletters
- Example: Classifying modification reports as adaptive, perfective or corrective based on documentation
  - Audris Mockus, Lawrence G. Votta: Identifying Reasons for Software Changes using Historic Databases. ICSM2000: pp. 120-130



## Archival Records

- Service records
  - Clients served over a period of time
- Organizational records
  - Organizational charts and budgets
- Layouts
  - Maps and charts
- Lists of names and relevant articles
- Survey data
  - Census records
- Personal records
  - Diaries, calendars, telephone lists
- Example: Study of parallel changes to source code was based on revision control logs
  - Dewayne E. Perry, Harvey P. Siy, Lawrence G. Votta: Parallel changes in large-scale software development: an observational case study. ACM TSE Methodology 10(3): 308-337 (2001)

## Interviews

- ❍ Open-ended interviews
  - Address facts and opinions about an event
  - Flexible structure of interview (or no structure at all!)
- ❍ Focused interviews
  - Short period of time (about an hour)
  - Similar approach as open-ended.
- ❍ Formal surveys
  - Produce quantifiable data
- ❍ Example: Used semi-structured interviews to understand the effect of distance on coordination in teams
  - Rebecca E. Grinter, James D. Herbsleb, Dewayne E. Perry: The geography of coordination: dealing with distance in R&D work. GROUP 1999: pp. 306-315

## Direct Observation

- ❍ Field visits- creates opportunity for direct observation
- ❍ Photographs of site
  - Need permission in order to proceed!
- ❍ Can be used to calibrate self-reports
  - Example: Informal, impromptu interactions
- ❍ Example: Followed software developers around to characterize how they spend their time
  - Dewayne E. Perry, Nancy A. Staudenmayer, Lawrence G. Votta: People, Organizations, and Process Improvement. IEEE Software 11(4): 36-45 (1994)

## Participant-observation

- ❍ Not a passive observer, but actually participate in setting
  - Employee of the company under study
- ❍ Provides an opportunity to understand the rationale and behavior of people and organization being studied
- ❍ Example: Seaman participated in 23 code inspections over period of five months at NASA/Goddard Space Flight Center's Flight Dynamics Division
  - Carolyn B. Seaman, Victor R. Basili: Communication and Organization: An Empirical Study of Discussion in Inspection Meetings. IEEE Trans. Software Eng. 24(7): 559-572 (1998)

## Physical Artifacts

- ❍ Technological tool, instrument, or device
- ❍ Artifacts can be collected or observed as part a field visit
- ❍ Works of art or types of physical evidence
- ❍ Example: Diachronic study of art records to determine whether right-handedness was a recent or old trait
  - Two rival hypotheses: Physiological predisposition vs Social/environmental pressure
  - Tested by counting unimanual tool usage in art representations
  - 1200 examples from 1500 BC to 1950, world sources
  - 92.6% used right hand
  - Geo/historical distribution uniformly high
  - Seems to support physiological interpretation that right-handedness is an age-old trait

## Sources of Evidence: Strengths, Weaknesses

| Source of Evidence | Strengths | Weaknesses |
|---|---|---|
| Documentation | ⁕ Stable - can be reviewed repeatedly<br>⁕ Unobtrusive - not created as a result of the case study<br>⁕ Exact - contains exact names, references, and details of an event<br>⁕ Broad coverage - long span of time, many events, and many settings | ⁕ Retrievability - can be low<br>⁕ Biased selectivity, if collection is incomplete<br>⁕ Reporting bias - reflects (unknown) bias of author<br>⁕ Access - may be deliberately blocked |
| Archival Records | {same as above for documentation}<br>⁕ Precise and quantitative | {same as above for documentation}<br>⁕ Accessibility due to privacy reasons |
| Interviews | ⁕ Targeted - focuses directly on case study topic<br>⁕ Insightful - provides perceived causal inferences | ⁕ Bias due to poorly constructed questions<br>⁕ Response bias<br>⁕ Inaccuracies due to poor recall<br>⁕ Reflexivity - interviewee gives what interview wants to hear |

Reproduced from Yin 2002

## Strengths and Weaknesses Cont'd

| Source of Evidence | Strengths | Weaknesses |
|---|---|---|
| Direct Observations | ⁕ Reality - covers events in real time<br>⁕ Contextual - covers content of event | ⁕ Time consuming<br>⁕ Selectivity - unless broad coverage<br>⁕ Reflexivity - event may proceed differently because it is being observed<br>⁕ Cost- hours needed by human observers |
| Participant Observations | {same as above for direct observation}<br>⁕ Insightful into interpersonal behavior and motives | {same as above for direct observation}<br>⁕ Bias due to investigator's manipulation of events |
| Physical Artifacts | ⁕ Insightful into cultural features<br>⁕ Insightful into technical operations | ⁕ Selectivity<br>⁕ Availability |

Reproduced from Yin 2002

## Principles of Data Collection

- ❍ Use Multiple Sources of Evidence
- ❍ Create a Case Study Database
- ❍ Maintain a Chain of Evidence

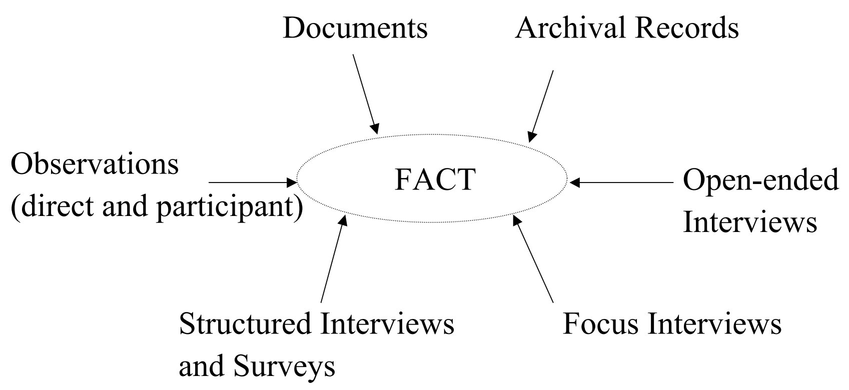
**These principles can be applied to all six data collection methods**

## Multiple Sources of Evidence

- ❍ Triangulation of data sources
- ❍ Basic idea: Collect evidence from more than one source pointing towards the same *facts*
  - **Warning: Collecting data from several sources *does not* guarantee data triangulation!**
- ❍ Example: Different approaches were used collect data about how developers spend their time.
  - Dewayne E. Perry, Nancy A. Staudenmayer, Lawrence G. Votta: People, Organizations, and Process Improvement. IEEE Software 11(4): 36-45 (1994)
    - Collected cross-sectional and direct observation data
  - Marc G. Bradac, Dewayne E. Perry, Lawrence G. Votta: Prototyping a Process Monitoring Experiment. IEEE TSE. 20(10): 774-784 (1994)
    - Collected longitudinal data

## Multiple Sources of Evidence

**Convergence of Evidence** **(Figure 4.2)**


Documents
Archival Records
Observations (direct and participant)
FACT
Open-ended Interviews
Structured Interviews and Surveys
Focus Interviews

## Case Study Database

- A case study database provides a formal assembly of evidence
- Elements to include:
  - Case study notes
    - From interviews, documents, etc.
    - Categorized, complete
  - Case study documents
    - Annotated bibliography of the documents- facilitates storage, retrieval, help future investigators share the database
  - Tabular materials
    - Survey and quantitative data
  - Narratives
    - Written by the researcher, providing tentative answers to the research questions
    - Connect pertinent issues

## Chain of Evidence

- ❍ Maintaining a chain of evidence is analogous to providing traceability in a software project
- ❍ Forms explicit links between
  - Questions asked
  - Data collected
  - Conclusion drawn

## Chain of Evidence

**Maintaining a Chain of Evidence**

$$\begin{tikzcd}
\text{Case Study Report} \arrow[d, leftrightarrow] \\
\text{Case Study Database} \arrow[d, leftrightarrow] \\
\text{Citations to Specific Evidentiary Sources in the Case Study D.B.} \arrow[d, leftrightarrow] \\
\text{Case Study Protocol (linking questions to protocol topics)} \arrow[d, leftrightarrow] \\
\text{Case Study Questions}
\end{tikzcd}$$

28th International Conference on Software Engineering

## Lunch

28th International Conference on Software Engineering

## Exercise: Design a Case Study

28th International Conference on Software Engineering

## Data Analysis and Validity

## Data Analysis

- ❍ Analytic Strategies
- ❍ 3 general strategies
  - Relying on theoretical propositions
  - Thinking about rival explanations
  - Developing a case description
- ❍ 5 specific analytic techniques
  - Pattern matching
  - Explanation building
  - Time-series analysis
  - Logic models
  - Cross-case synthesis
- ❍ Criteria for high quality analysis

## Characteristics of Case Study Analysis

- We have the data. Now what?
  - Don't collect the data if you don't know how to analyze them!
  - Figure out your analysis strategies when designing the case study
- Data analysis should use the evidence to address the propositions
  - Both quantitative and qualitative evidence should be analyzed
- Difficult step because strategies and techniques have not been well defined
  - Big stumbling block of many case studies

## General Strategy
## 1. Relying on Theoretical Propositions

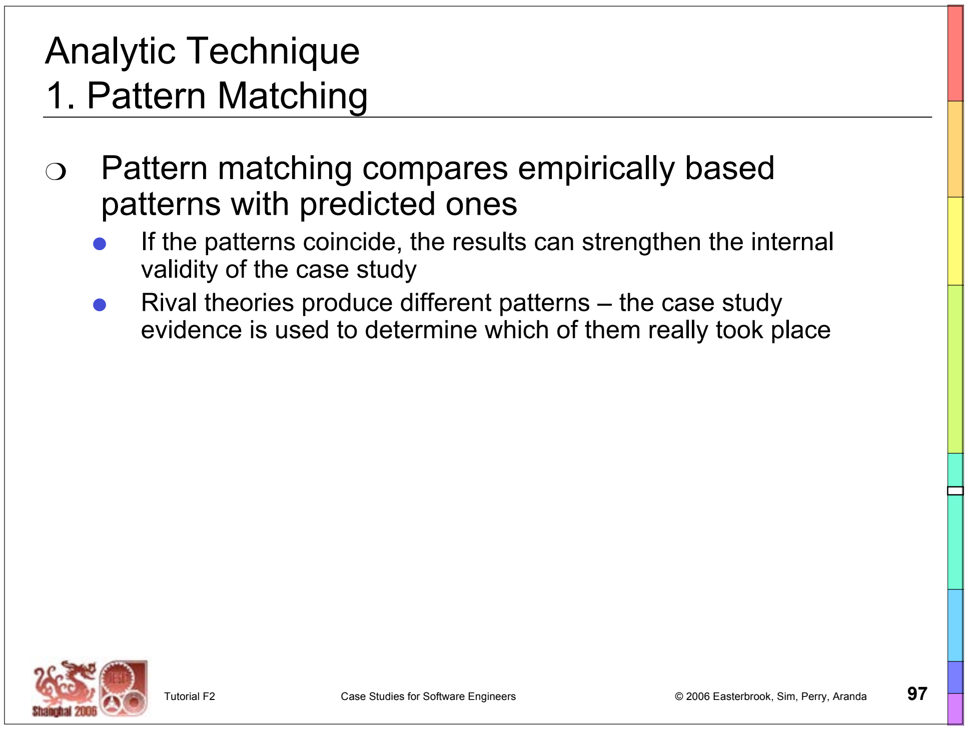
- Shapes the data collection plan and prioritizes the relevant analytic strategies
- Helps to focus attention on theory-relevant data and to ignore irrelevant data
- Helps to organize the entire case study and define rival explanations to be examined

## General Strategy
## 2. Thinking About Rival Explanations

- Especially useful when doing case study evaluations
- Attempts to collect evidence about other possible influences
- The more rivals the analysis addresses and rejects, the more confidence can be placed in the findings

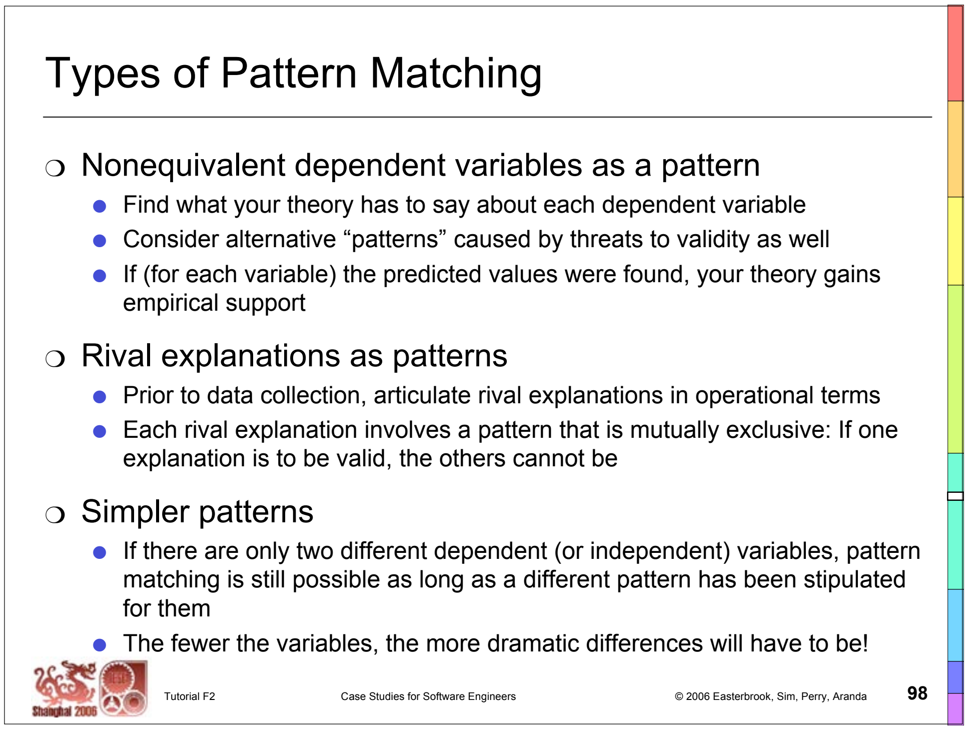
## General Strategy
## 3. Developing a Case Description

- Serves as an alternative when theoretical proposition and rival explanation are not applicable
- A descriptive approach may help to identify the appropriate causal links to be analyzed



## Analytic Technique
## 1. Pattern Matching

- ❍ Pattern matching compares empirically based patterns with predicted ones
  - If the patterns coincide, the results can strengthen the internal validity of the case study
  - Rival theories produce different patterns – the case study evidence is used to determine which of them really took place



## Types of Pattern Matching

- ❍ Nonequivalent dependent variables as a pattern
  - Find what your theory has to say about each dependent variable
  - Consider alternative “patterns” caused by threats to validity as well
  - If (for each variable) the predicted values were found, your theory gains empirical support
- ❍ Rival explanations as patterns
  - Prior to data collection, articulate rival explanations in operational terms
  - Each rival explanation involves a pattern that is mutually exclusive: If one explanation is to be valid, the others cannot be
- ❍ Simpler patterns
  - If there are only two different dependent (or independent) variables, pattern matching is still possible as long as a different pattern has been stipulated for them
  - The fewer the variables, the more dramatic differences will have to be!

## Analysis Technique
## 2. Explanation Building

- ❍ Special type of pattern matching
- ❍ Stipulate a presumed set of causal links
- ❍ Series of iterations in building explanation
    1. Make initial theoretical statement
    2. Compare the findings of the initial case against statement
    3. Revise the statement
    4. Compare other details of the case against the revision
    5. Compare the revisions to the facts of 2nd, 3rd or more cases
    6. Repeat the process if needed
- ❍ Difficult, dangerous technique!
    - You may drift away from your original focus
    - Need much analytic insight and emphasis in considering all evidence

## Analysis Techniques
## 3. Time Series Analysis

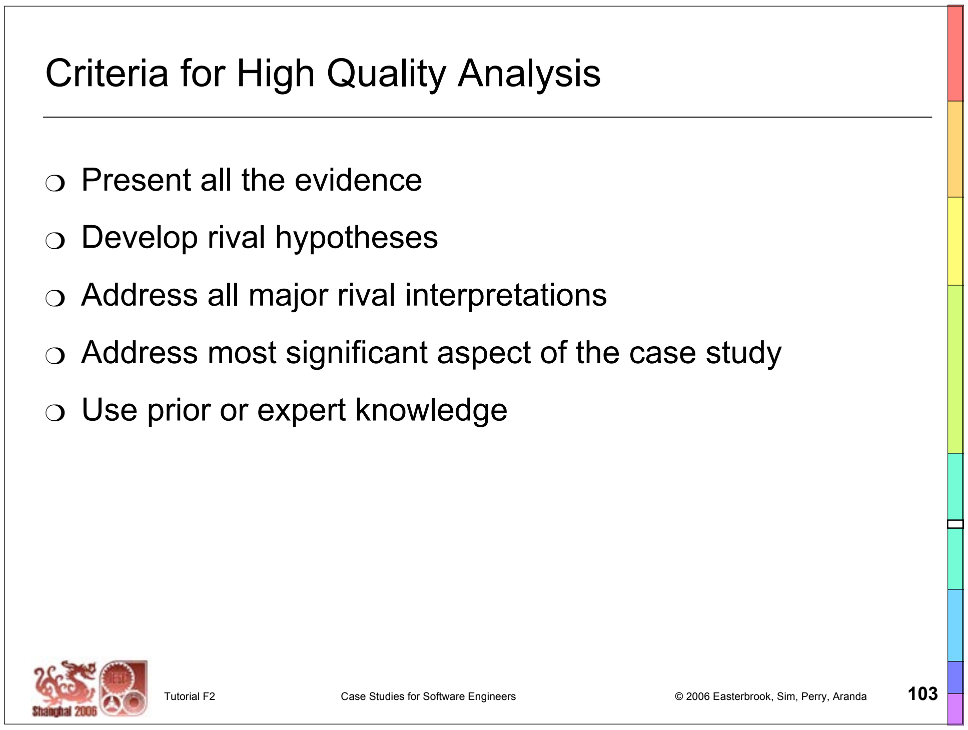
- ❍ Theory and analysis are concerned with events over time
- ❍ Types of Time Series Analyses:
    - Simple Time Series
        - Single variable
        - Statistical tests are feasible
        - Compare the tendencies found against the expected tendency, rival-theory tendencies, and tendencies expected due to threats to validity
    - Complex Time Series
        - Multiple sets of relevant variables
        - Statistical tests no longer feasible, but many patterns to observe
    - Chronologies
        - Some events must always occur *before* others
        - Some events must always be *followed* by others
        - Some events can only follow others after a specified *interval of time*
        - Certain *time periods* in a case study have classes of events substantially different from those of other periods

## Analysis Technique
## 4. Logic Models

- Deliberately stipulate a complex chain of events over time
  - Events are staged in repeated cause-effect-cause-effect patterns
  - Similar to pattern matching, but focus is on sequential stages
  - Logic models are analogous to other conceptual models familiar to software engineers (SADT, statecharts, etc.)
- Four types of logic models:
  - Individual-Level Logic Model
    - Assumes the case study is about a single person
  - Firm or Organizational-Level Logic Model
    - Traces events taking place in an organization
  - Alternative configuration for an Organizational-Level Logic Model
    - Emphasize systemic changes that are not necessarily linear (e.g. reforming or transforming an organization)
  - Program-Level Logic Model
    - Analyzes the structure of a program and its outcomes

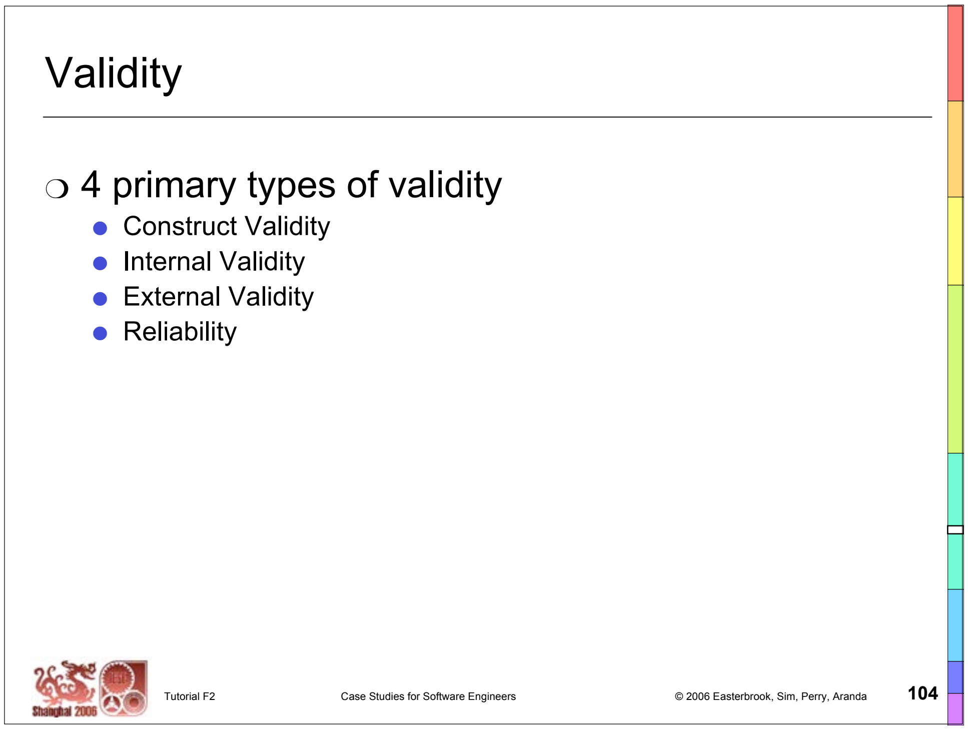
## Analysis Technique
## 5. Cross-Case Synthesis

- A research synthesis technique
- Guidelines
  - Treat each individual case study as a separate study
    - Just as you would treat different experiments separately
  - Create word tables that display data, in a uniform framework, from each case
  - Examine word tables for cross-case patterns
  - You will need to rely strongly on argumentative interpretation, not numeric properties
    - Numeric properties cannot be pooled since each case is different



## Criteria for High Quality Analysis

- ❍ Present all the evidence
- ❍ Develop rival hypotheses
- ❍ Address all major rival interpretations
- ❍ Address most significant aspect of the case study
- ❍ Use prior or expert knowledge



## Validity

- ❍ 4 primary types of validity
  - ● Construct Validity
  - ● Internal Validity
  - ● External Validity
  - ● Reliability

## Construct Validity

- ❍ Are we measuring what we intend to measure?
  - Akin to the requirements problem: are we building the right system?
  - If we don't get this right, the rest doesn't matter
- ❍ Constructs: abstract concepts
  - Theoretical constructions
  - Must be operationalized in the experiment
- ❍ Necessary condition for successful experiment
- ❍ Divide construct validity into three parts:
  - Intentional Validity
  - Representation Validity
  - Observation Validity

## Construct Validity

- ❍ Intentional Validity
  - Do the constructs we chose adequately represent what we intend to study
  - Akin to the requirements problem where our intent is fair scheduling but our requirement is FIFO
  - Are our constructs specific enough?
  - Do they focus in the right direction?
  - Eg, is it intelligence or cunningness?

## Construct Validity

- ❍ Representation Validity
  - How well do the constructs or abstractions translate into observable measures
  - Two primary questions:
    - Do the sub-constructs properly define the constructs?
    - Do the observations properly interpret, measure or test the constructs?
- ❍ 2 ways to argue for representation validity
  - Face validity (that is, “looks appropriate”)
    - Very weak argument!
    - Strengthened by consensus of experts
  - Content validity
    - Check the operationalization against the domain for the construct
    - The extent to which the tests measure the content of the domain being tested - ie, cover the domain
  - Both are qualitative judgments

## Construct Validity

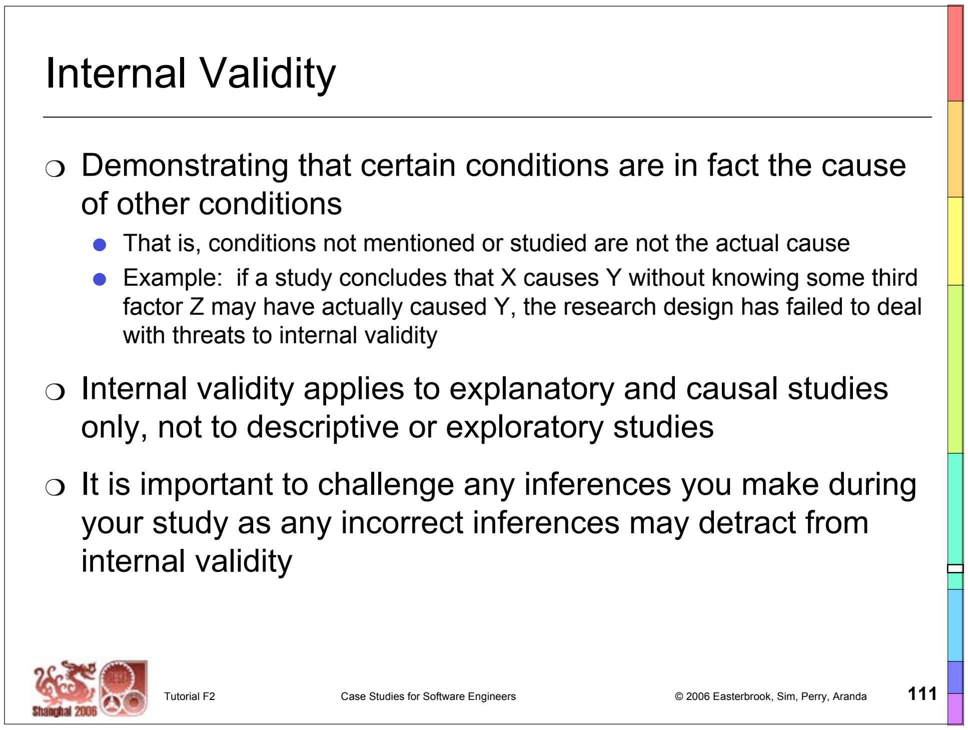
- ❍ Observation Validity
  - How good are the measures themselves?
  - Different aspects illuminated by
    - Predictive validity
    - Criterion validity
    - Concurrent validity
    - Convergent validity
    - Discriminant validity
- ❍ Predictive Validity
  - Observed measure predicts what it should predict and nothing else
    - E.g., college aptitude tests are assessed for their ability to predict success in college
- ❍ Criterion Validity
  - Degree to which the results of a measure agree with those of an independent standard
    - Eg, for college aptitude, GPA or successful first year

## Construct Validity

- Concurrent Validity
  - The observed measure correlates highly with an established set of measures
    - Eg, shorter forms of tests against longer forms
- Convergent Validity
  - Observed measure correlates highly with other observable measures for the same construct
  - Utility is not that it duplicates a measure but is a new way of distinguishing a particular trait while correlating with similar measures
- Discriminant Validity
  - The observable measure distinguishes between two groups that differ on the trait in question
  - Lack of divergence argues for poor discriminant validity

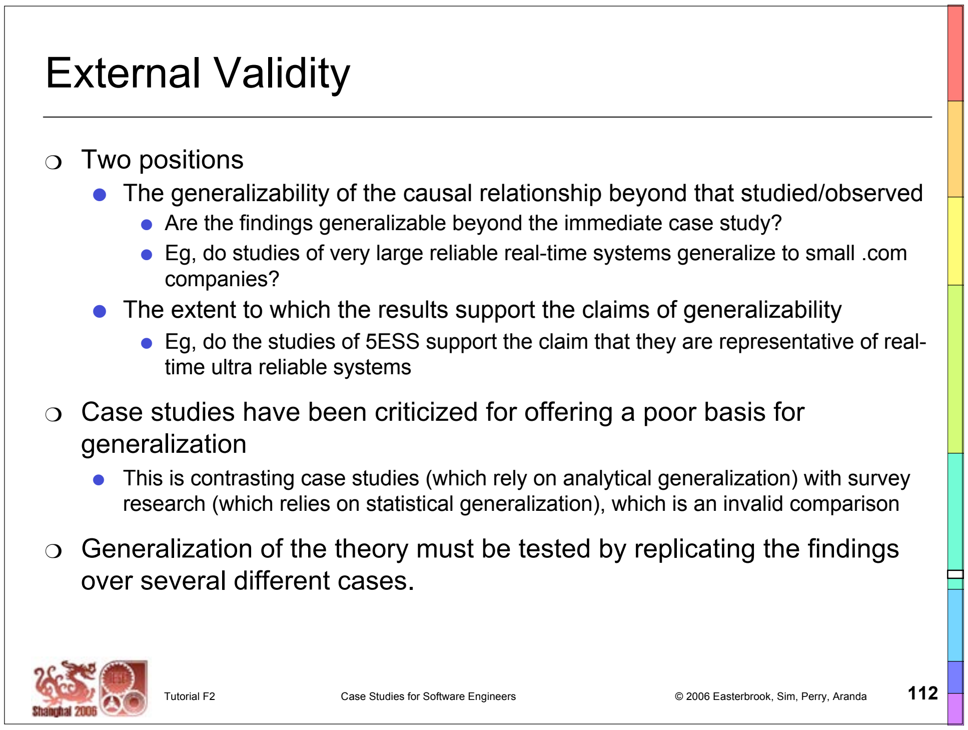
## Internal Validity

- Can we be sure our results really follow from the data?
  - Have we adequately ruled out rival hypotheses?
- Have we eliminated confounding variables?
  - Participant variables
  - Researcher variables
  - Stimulus, procedural and situational variables
  - Instrumentation
  - Nuisance variables
- Confounding sources of internal invalidity
  - H: History
    - events happen during the study (eg, 9/11)
  - M: Maturation
    - older/wiser/better between during study
  - I: Instrumentation
    - change due to observation/measurement instruments
  - S: Selection
    - differing nature of participants
    - effects of choosing participants



## Internal Validity

- Demonstrating that certain conditions are in fact the cause of other conditions
  - That is, conditions not mentioned or studied are not the actual cause
  - Example: if a study concludes that X causes Y without knowing some third factor Z may have actually caused Y, the research design has failed to deal with threats to internal validity
- Internal validity applies to explanatory and causal studies only, not to descriptive or exploratory studies
- It is important to challenge any inferences you make during your study as any incorrect inferences may detract from internal validity



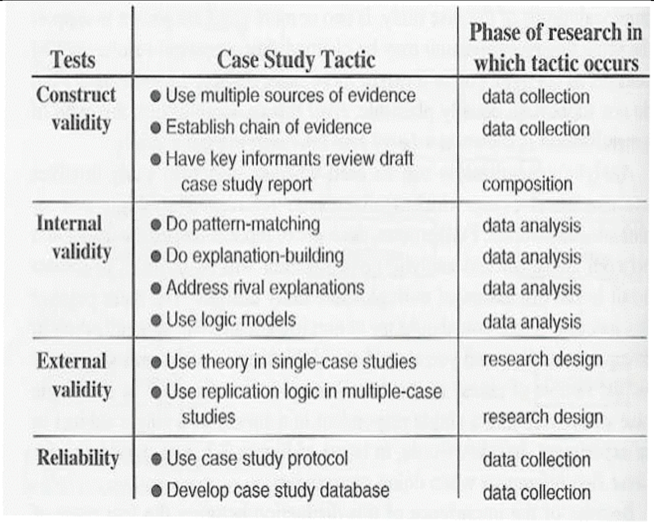
## External Validity

- Two positions
  - The generalizability of the causal relationship beyond that studied/observed
    - Are the findings generalizable beyond the immediate case study?
    - Eg, do studies of very large reliable real-time systems generalize to small .com companies?
  - The extent to which the results support the claims of generalizability
    - Eg, do the studies of 5ESS support the claim that they are representative of real-time ultra reliable systems
- Case studies have been criticized for offering a poor basis for generalization
  - This is contrasting case studies (which rely on analytical generalization) with survey research (which relies on statistical generalization), which is an invalid comparison
- Generalization of the theory must be tested by replicating the findings over several different cases.

## Reliability

- ❍ Demonstrating that the operations of a study can be repeated with the same results
  - ● Note: the repetition of the study should occur on the same case, not "replicating" the results on a different case
- ❍ "The goal of reliability is to minimize the errors and biases in a study"
- ❍ A prerequisite for reliability testing is documented procedures for the case study
- ❍ A good guideline is to perform research so that an auditor could follow the documented procedures and arrive at the same results

## Tactics to Address Quality in Case Study Design



| Tests | Case Study Tactic | Phase of research in which tactic occurs |
|---|---|---|
| Construct validity | ● Use multiple sources of evidence | data collection |
| | ● Establish chain of evidence | data collection |
| | ● Have key informants review draft case study report | composition |
| Internal validity | ● Do pattern-matching | data analysis |
| | ● Do explanation-building | data analysis |
| | ● Address rival explanations | data analysis |
| | ● Use logic models | data analysis |
| External validity | ● Use theory in single-case studies | research design |
| | ● Use replication logic in multiple-case studies | research design |
| Reliability | ● Use case study protocol | data collection |
| | ● Develop case study database | data collection |

Case Study Tactics for the Four Design Tests (Yin, page 34)

28th International Conference on Software Engineering

## Exercise: Design a Case Study (concludes)

28th International Conference on Software Engineering

## Publishing Case Studies

## Overview of this Section

❍ Targeting Case Study Reports

❍ Composition Styles for Case Studies

❍ General Guidelines from Software Engineering

❍ Sample Paper Outlines

## Targeting Case Study Reports

❍ more diverse audiences than other research methods
- no single report will serve all audiences simultaneously
- may need more than one version of a case study report

❍ A case study report is a significant communication device
- … can communicate deep insights about a phenomenon to a variety of non-specialists
- Practitioners like case studies, so context is important

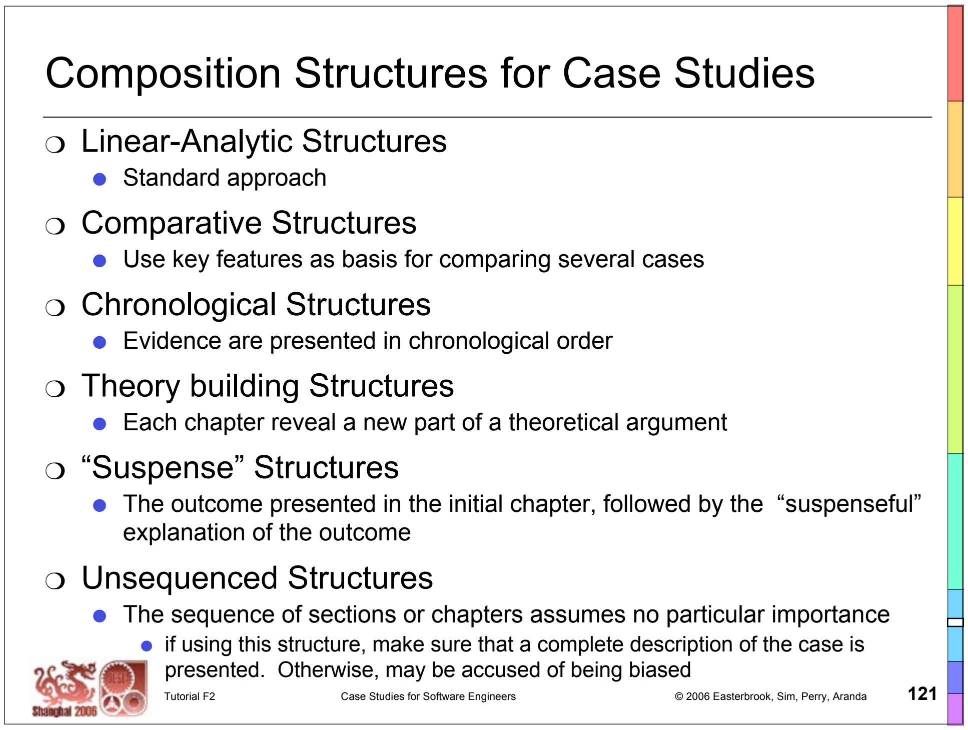
❍ Orient the case study report to an audience
- preferences of the potential audience should dictate the form of your case study report
- Greatest error is to compose a report from an egocentric perspective
- Therefore: identify the audience before writing a case study report
- Recommendation: examine previous case study reports that have successfully communicated with the identified audience

## Formats for Written Case Study Reports

- Classic single-case study
  - - a single narrative is used to describe and analyze the case
- Multiple-case version of this classic single case
  - - individual cases are presented in separate chapters
  - - also contain chapters that contain cross-case analysis
- Non-traditional narrative (single or multiple)
  - - use question-and-answer format
- Cross-case analysis (multiple-case studies only)
  - - entire report consist of cross-case analysis
  - - each chapter would be devoted to a separate cross-case issue
- Note: Format should be identified during the design phase of case study

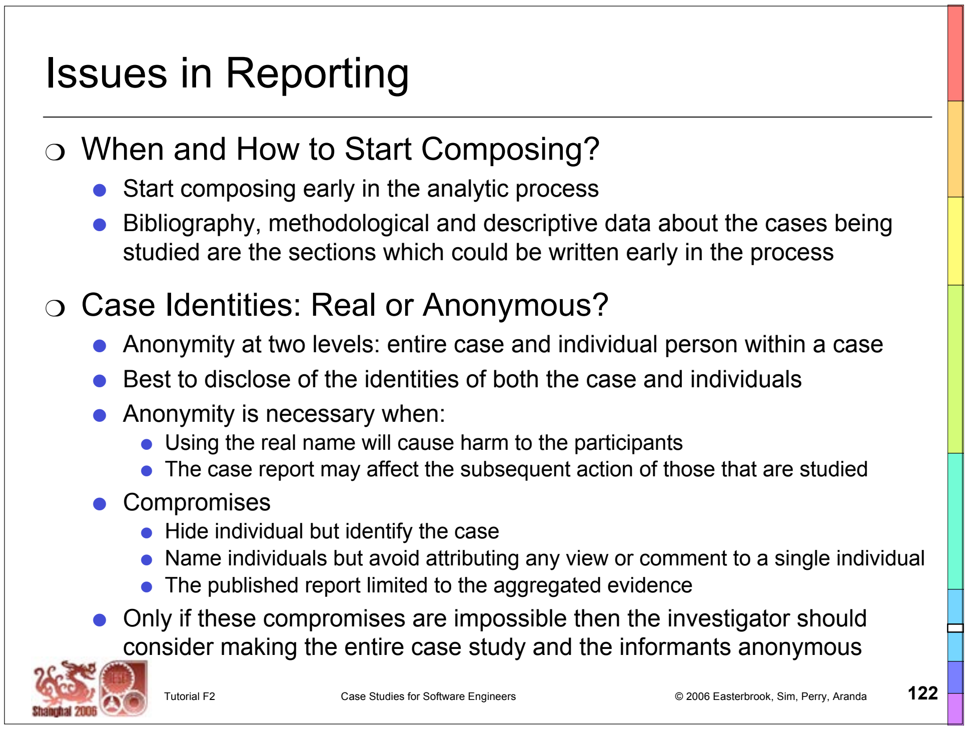
## Sequences of Studies

- Empirical studies should be performed as part of a sequence
  - Each going deeper or shedding light on a different aspect of a problem
  - Can deploy different tactics, strategies, methods
- Rationales to use case study as part of a larger, multimethod study:
  1. To determine whether converging evidence might be obtained even though different methods have been used
  2. After analyzing data collected by other methods, case study might be able to illustrate an individual case in greater depth
  3. Case study may be used to elucidate some underlying process which another method is used to define the prevalence or frequency of such processes



## Composition Structures for Case Studies

- ❍ Linear-Analytic Structures
  - Standard approach
- ❍ Comparative Structures
  - Use key features as basis for comparing several cases
- ❍ Chronological Structures
  - Evidence are presented in chronological order
- ❍ Theory building Structures
  - Each chapter reveal a new part of a theoretical argument
- ❍ “Suspense” Structures
  - The outcome presented in the initial chapter, followed by the “suspenseful” explanation of the outcome
- ❍ Unsequenced Structures
  - The sequence of sections or chapters assumes no particular importance
    - if using this structure, make sure that a complete description of the case is presented. Otherwise, may be accused of being biased



## Issues in Reporting

- ❍ When and How to Start Composing?
  - Start composing early in the analytic process
  - Bibliography, methodological and descriptive data about the cases being studied are the sections which could be written early in the process
- ❍ Case Identities: Real or Anonymous?
  - Anonymity at two levels: entire case and individual person within a case
  - Best to disclose of the identities of both the case and individuals
  - Anonymity is necessary when:
    - Using the real name will cause harm to the participants
    - The case report may affect the subsequent action of those that are studied
  - Compromises
    - Hide individual but identify the case
    - Name individuals but avoid attributing any view or comment to a single individual
    - The published report limited to the aggregated evidence
  - Only if these compromises are impossible then the investigator should consider making the entire case study and the informants anonymous

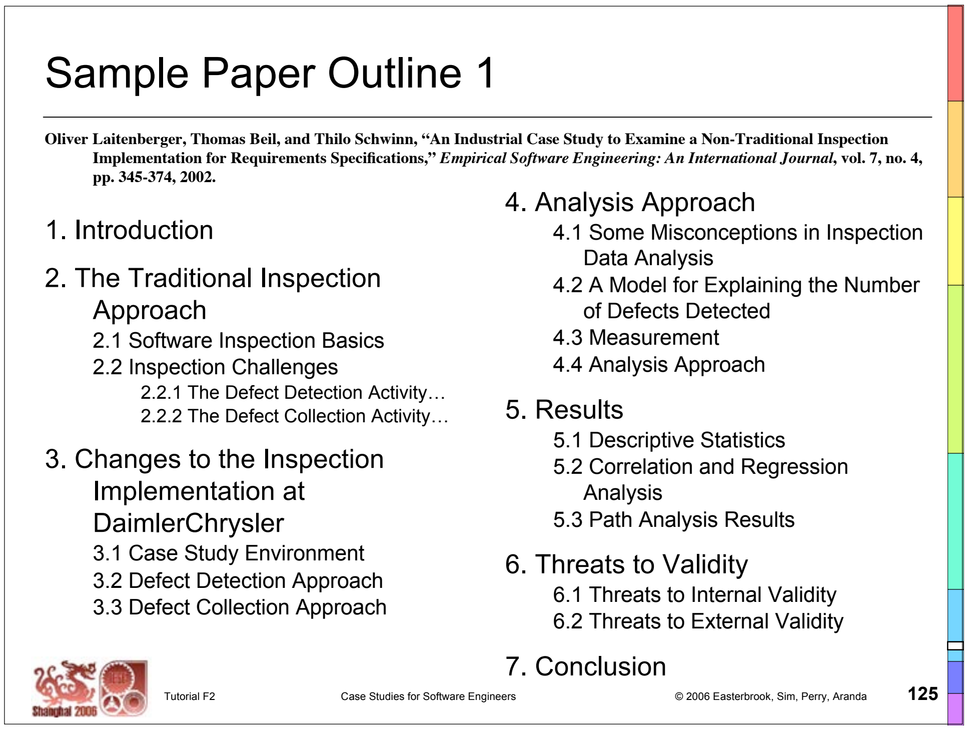
## Issues in Reporting

- ❍ Review of the Draft Case Study: A Validating Procedure
  - There should be a draft report
  - It should be reviewed by peers, and by the participants and informants in the case
  - The reviewers may disagree with the conclusion and interpretations, but not the actual facts of the case
  - This process increases the construct validity of the study and reduced the likelihood of falsely reporting an event

## General Guidelines from SE

**Barbara A. Kitchenham, Shari Lawrence Pfleeger, Lesley M. Pickard, Peter W. Jones, David C. Hoaglin, Khaled El Emam, and Jarrett Rosenberg, "Preliminary Guidelines for Empirical Research in Software Engineering," IEEE Transactions on Software Engineering, Vol. 28, No 8, August 2002.**

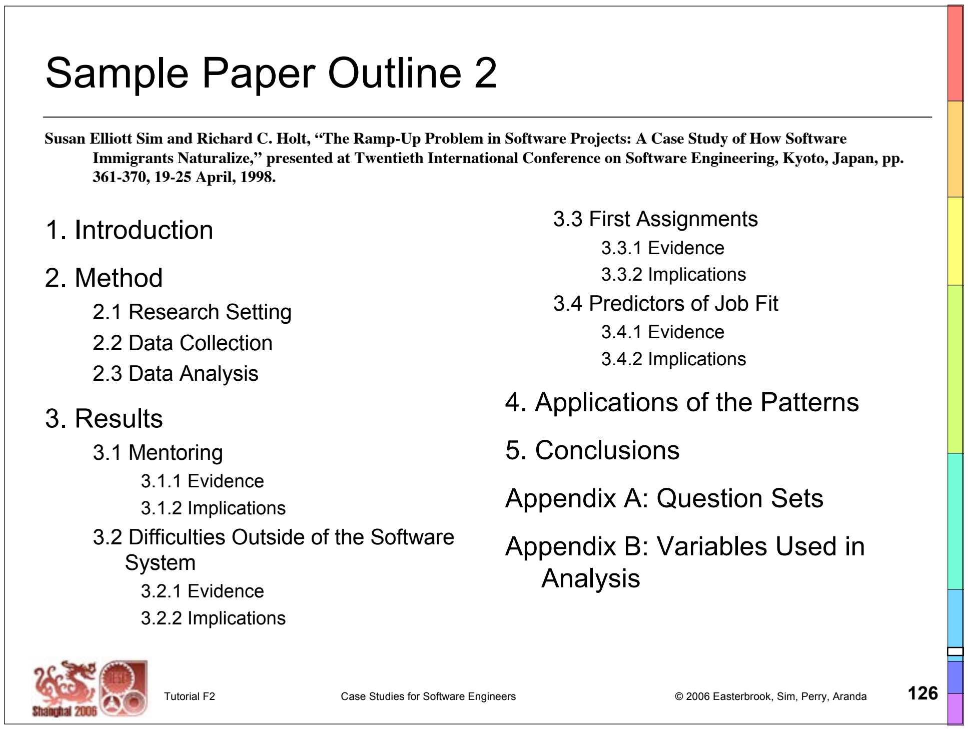
- ❍ Empirical Context
- ❍ Study Design
- ❍ Conducing the Case Study and Data Collection
- ❍ Analysis
- ❍ Presentation of Results
- ❍ Interpretation of Results



## Sample Paper Outline 1

**Oliver Laitenberger, Thomas Beil, and Thilo Schwinn, "An Industrial Case Study to Examine a Non-Traditional Inspection Implementation for Requirements Specifications,"** ***Empirical Software Engineering: An International Journal*****, vol. 7, no. 4, pp. 345-374, 2002.**

1. Introduction
2. The Traditional Inspection Approach
    - 2.1 Software Inspection Basics
    - 2.2 Inspection Challenges
        - 2.2.1 The Defect Detection Activity…
        - 2.2.2 The Defect Collection Activity…
3. Changes to the Inspection Implementation at DaimlerChrysler
    - 3.1 Case Study Environment
    - 3.2 Defect Detection Approach
    - 3.3 Defect Collection Approach
4. Analysis Approach
    - 4.1 Some Misconceptions in Inspection Data Analysis
    - 4.2 A Model for Explaining the Number of Defects Detected
    - 4.3 Measurement
    - 4.4 Analysis Approach
5. Results
    - 5.1 Descriptive Statistics
    - 5.2 Correlation and Regression Analysis
    - 5.3 Path Analysis Results
6. Threats to Validity
    - 6.1 Threats to Internal Validity
    - 6.2 Threats to External Validity
7. Conclusion



## Sample Paper Outline 2

**Susan Elliott Sim and Richard C. Holt, "The Ramp-Up Problem in Software Projects: A Case Study of How Software Immigrants Naturalize," presented at Twentieth International Conference on Software Engineering, Kyoto, Japan, pp. 361-370, 19-25 April, 1998.**

1. Introduction
2. Method
    - 2.1 Research Setting
    - 2.2 Data Collection
    - 2.3 Data Analysis
3. Results
    - 3.1 Mentoring
        - 3.1.1 Evidence
        - 3.1.2 Implications
    - 3.2 Difficulties Outside of the Software System
        - 3.2.1 Evidence
        - 3.2.2 Implications
    - 3.3 First Assignments
        - 3.3.1 Evidence
        - 3.3.2 Implications
    - 3.4 Predictors of Job Fit
        - 3.4.1 Evidence
        - 3.4.2 Implications
4. Applications of the Patterns
5. Conclusions

Appendix A: Question Sets

Appendix B: Variables Used in Analysis

## Sample Paper Outline 3

**Dewayne E. Perry, Harvey P. Siy, and Lawrence G. Votta, "Parallel Changes in Large Scale Software Development: An Observational Case Study," presented at Twentieth International Conference on Software Engineering, pp. 251-260, 19-25 April 1998.**

1. Introduction
2. Related Work
   - 2.1 Configuration Management
   - 2.2 Program Analysis
   - 2.3 Build Coordination
   - 2.4 Empirical Evaluation
3. Study Context
   - 3.1 The System Under Study
   - 3.2 The 5ESS Change Management Process
4. Data and Analysis
   - 4.1 Levels of Parallel Development
   - 4.2 Effects of Parallel Development on a File
   - 4.3 Interfering Changes
   - 4.4 Multilevel Analysis of Parallel Development
   - 4.5 Parallel Versions
5. Validity
6. Summary and Evaluation
   - 6.1 Study Summary
   - 6.2 Evaluation of Current Support
   - 6.3 Future Directions

28th International Conference on Software Engineering

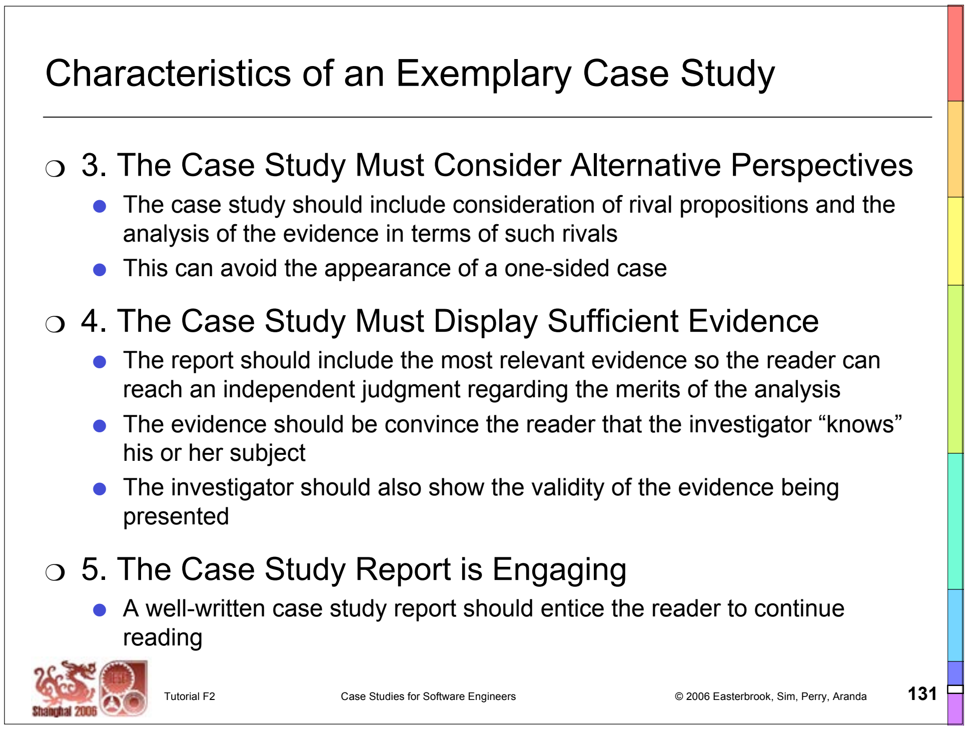
## Reviewing Case Studies

## What Makes an Exemplary Case Study?

- The exemplary case study goes beyond the methodological procedures
- Mastering the techniques does not guarantee an exemplary case study

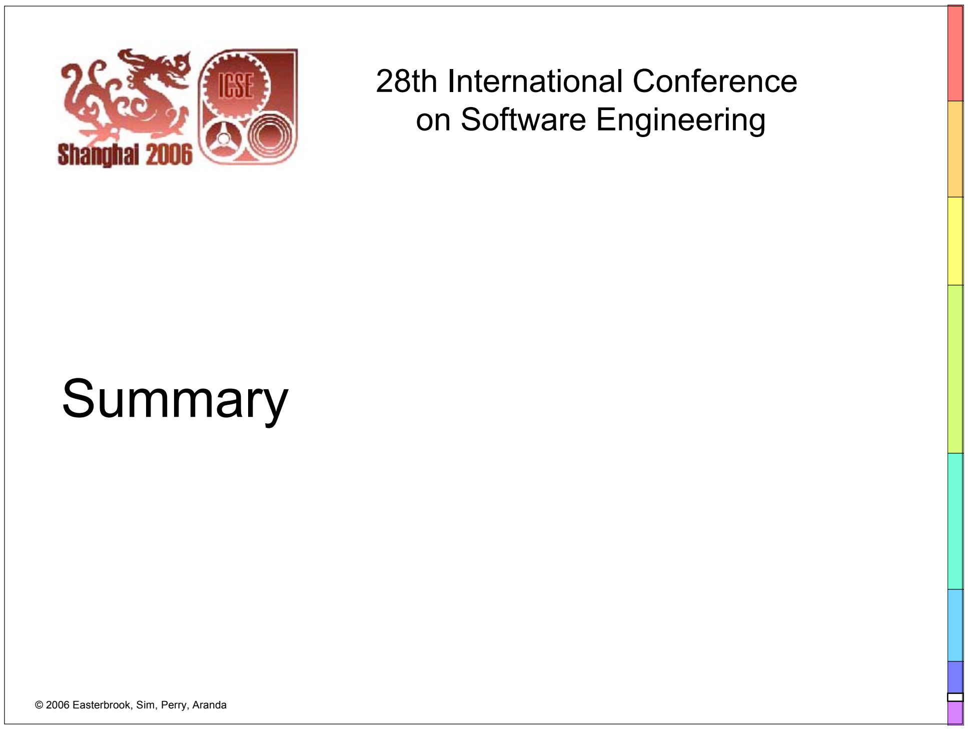
## Characteristics of an Exemplary Case Study

- 1. The Case Study Must Be Significant
  - The case should be unusual and of general public interest
  - The issue are important, either in theory or practical terms
    - Relevant to scientific understanding or to policy decisions
  - Prior to selecting a case study, the contribution should be described in detail assuming that the intended case study were to be completed successfully
- 2. The Case Study Must be “Complete”
  - Completeness can be characterized in at least three ways:
    - The boundaries of the case are given explicit attention
    - Exhaustive effort is spent on collecting all the relevant evidence
    - The case study was not ended because of nonresearch constraints



## Characteristics of an Exemplary Case Study

- 3. The Case Study Must Consider Alternative Perspectives
  - The case study should include consideration of rival propositions and the analysis of the evidence in terms of such rivals
  - This can avoid the appearance of a one-sided case
- 4. The Case Study Must Display Sufficient Evidence
  - The report should include the most relevant evidence so the reader can reach an independent judgment regarding the merits of the analysis
  - The evidence should be convince the reader that the investigator “knows” his or her subject
  - The investigator should also show the validity of the evidence being presented
- 5. The Case Study Report is Engaging
  - A well-written case study report should entice the reader to continue reading



28th International Conference on Software Engineering

## Summary

## Case Study as a Research Method

- ❍ The case study is a distinct research method
  - Has its own research designs
  - It is not a subset or variant of research designs used for other strategies
- ❍ Scientific
  - Synergistic relationship between theory and data
  - Starting a case study requires a theoretical orientation, which drives data collection
- ❍ Useful for answering "how" and "why" questions
  - In contrast to who, what, when, how many, how much
  - How, why = explanatory, descriptive
- ❍ Does not require control over events
  - More observational
- ❍ Focus on contemporary events
  - Less historical

## Key Message

- ❍ A case study is an empirical inquiry that
  - Investigates a contemporary phenomenon within its real-life context, especially when
  - The boundaries between phenomenon and context are not clearly evident.
- ❍ The case study inquiry
  - Copes with the technically distinctive situation in which there will be many more variables of interest that data points,
  - Relies on multiple sources of evidence, with data needing to converge in a triangulating fashion,
  - Benefits from the prior development of theoretical propositions to guide data collection and analysis.

## Further Reading

- Yin, R. K. (2002) Case Study Research: Design and Methods (3rd Edition). CA:Sage.
- Stake, R.E. (1995). The art of case study research. Thousand Oaks, CA:Sage.
- Ragin, C.C., & Becker, H.S. (Eds.). (1992). What is  a case? Exploring the foundations of social inquiry. Cambridge, UK: Cambridge University Press.