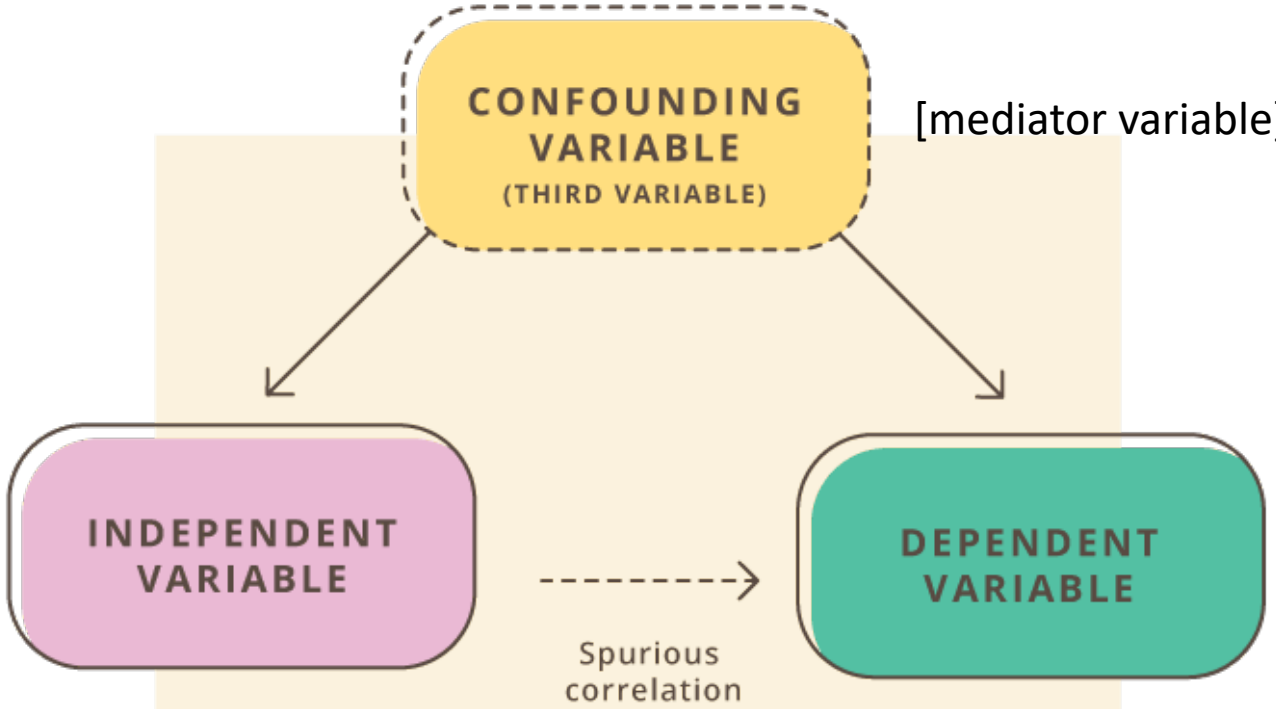
ECE1724H S2: Empirical Software Engineering

Case Studies & Ethics

The Edward S. Rogers Sr. Department
of Electrical & Computer Engineering
UNIVERSITY OF TORONTO



[mediator variable]

What at first looks like a causal relationship between IV and DV is ultimately spurious. The confounding variable is the hidden explanation.

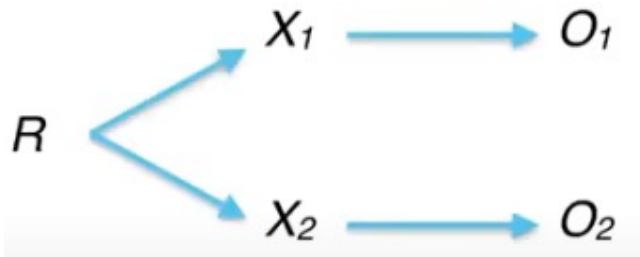
https://explorable.com/confounding-variables

The Vocabulary of Experiments

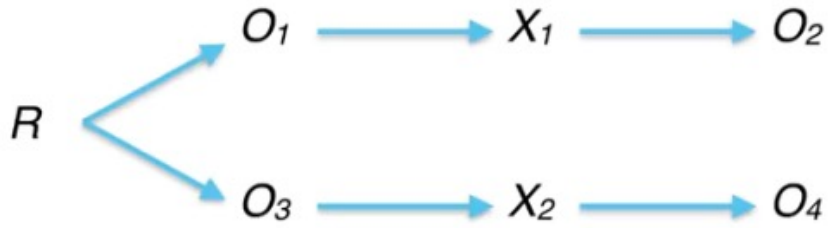
- **Experiment**: A study in which an intervention is deliberately introduced to observe its effects.
- **Randomized Experiment**: An experiment in which units are assigned to receive the treatment or an alternative condition by a random process such as the toss of a coin or a table of random numbers.
- **Quasi-Experiment**: An experiment in which units are not assigned to conditions randomly.
- **Natural Experiment**: Not really an experiment because the cause usually cannot be manipulated; a study that contrasts a naturally occurring event such as an earthquake with a comparison condition.
- **Correlational Study**: Usually synonymous with nonexperimental or observational study; a study that simply observes the size and direction of a relationship among variables.

Randomized Experiment Design

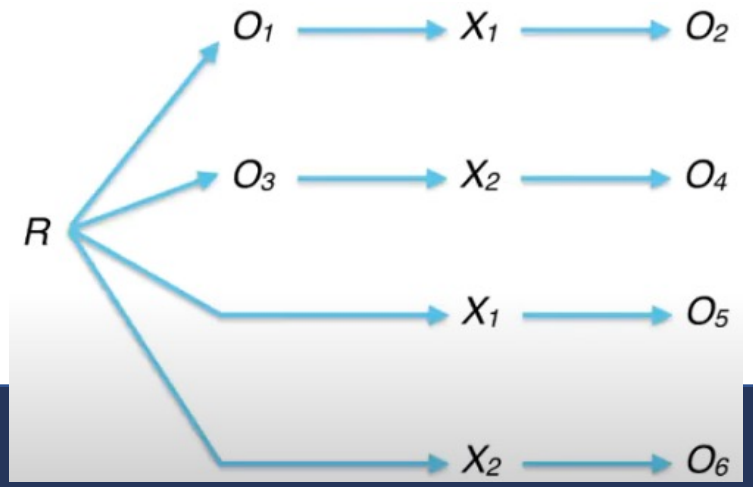
- Randomized two-group design



- Pretest-Posttest two-group design



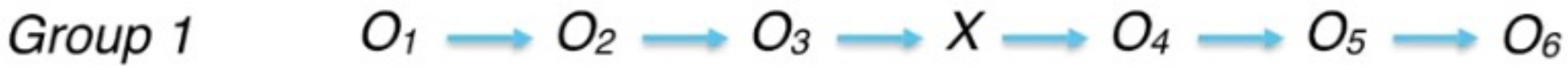
- Solomon four-group design



Quasi-experiments Design

- One-group Pretest-Posttest design

- Interrupted time series design



Typical Problems

- Construct Validity
  - Using things that are easy to measure instead of the intended concept
  - Wrong scale; insufficient discriminatory power
- Internal Validity
  - Confounding variables: Familiarity and learning;
  - Unmeasured variables: time to complete task, quality of result, etc.
- External Validity
  - Task representativeness: toy problem?
  - Subject representativeness: students for professional developers!
- Theoretical Reliability
  - Researcher bias: subjects know what outcome you prefer

Agenda for Today

- Paper reading presentation
- Case studies
- Ethical consideration

Paper reading

1(1). Sobel, Ann E. Kelley, and Michael R. Clarkson. "Formal methods application: An empirical tale of software development." *IEEE Transactions on Software Engineering* 28.3 (2002): 308-320.

1(2). Berry, Daniel M., and Walter F. Tichy. "Comments on" Formal methods application: an empirical tale of software development"." *IEEE Transactions on Software Engineering* 29.6 (2003): 567-571.

1(3). Sobel, Ann E. Kelley, and Michael R. Clarkson. "Response to" Comments on'Formal methods application: an empirical tale of software development"'." *IEEE Transactions on Software Engineering* 29.6 (2003): 572-575.

2. Lim, Sook. "How and why do college students use Wikipedia?." *Journal of the American Society for Information science and Technology* 60.11 (2009): 2189-2202.

Agenda for Today

- Paper reading presentation
- Case studies
- Ethical consideration


CASE STUDY

What is Case Study Research?

- The case study is a rigorous research approach or strategy that facilitates exploration of a contemporary phenomenon (i.e. “case”) in depth within its context using a variety of data sources.
- This ensures that the issue is not explored through one lens but rather a variety of lenses which allows for multiple facets of the phenomenon to be revealed and understood.
- Case studies *may* contain quantitative AND qualitative design components.

References: Pamela Baxter & Susan Jack (2008) Robert K. Yin (2009)

When to use case studies?

	(1)	(2)	(3)
METHOD	Form of Research Question	Requires Control of Behavioral Events?	Focuses on Contemporary Events?
Experiment	how, why?	yes	yes
Survey	who, what, where, how many, how much?	no	yes
Archival Analysis	who, what, where, how many, how much?	no	yes/no
History	how, why?	no	no
Case Study	how, why?	no	yes

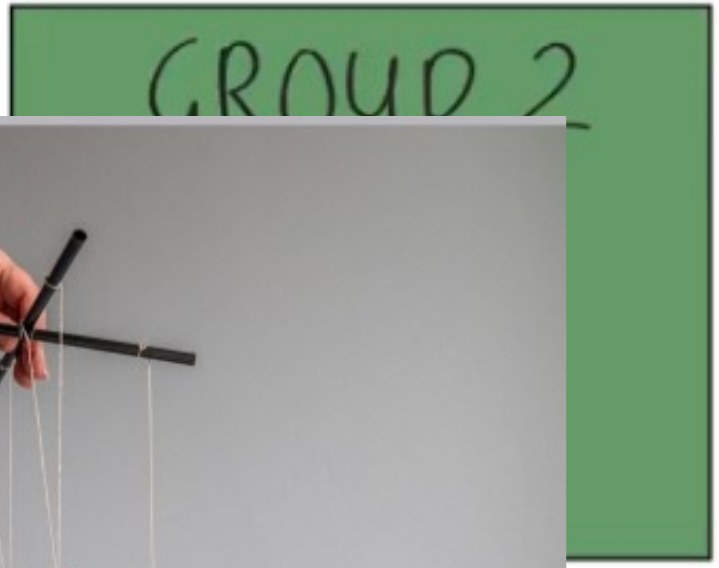
Figure 1.1 Relevant Situations for Different Research Methods SOURCE: COSMOS Corporation (1983)

When should you use a case study?

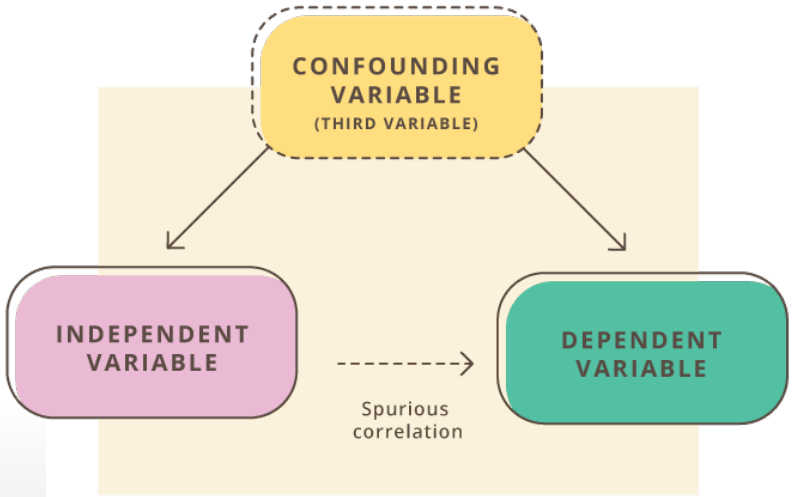
- When you can't control the variables
- When there are many more variables than data points
- When you cannot separate phenomena from context
  - Phenomena that don't occur in a lab setting
  - E.g. large scale, complex software projects
  - Effects can be wide-ranging.
  - Effects can take a long time 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


Randomised Experiment (comparing 2 treatments)
DATA-insight
A


GROUP 1


GROUP 2


Just a reminder...


CONFOUNDING VARIABLE
(THIRD VARIABLE)
INDEPENDENT VARIABLE
DEPENDENT VARIABLE
Spurious correlation

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

Do Developers Discover New Tools On The Toilet?

Emerson Murphy-Hill
Google, LLC
emersonm@google.com

Edward K. Smith*
Bloomberg
esmith404@bloomberg.net

Caitlin Sadowski
Google, LLC
supertri@google.com

Ciera Jaspan
Google, LLC
ciera@google.com

Collin Winter*
Waymo
collinwinter@waymo.com

Matthew Jorde
Google, LLC
majorde@google.com

Andrea Knight
Google, LLC
aknight@google.com

Andrew T
Google, L
atrenk@goog

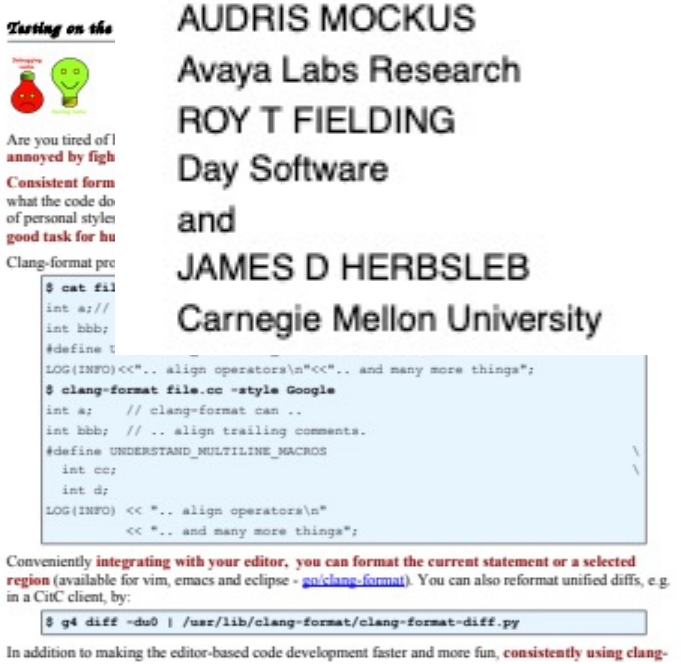
***Abstract*—Maintaining awareness of useful tools is a substantial challenge for developers. Physical newsletters are a simple technique to inform developers about tools. In this paper, we evaluate such a technique, called Testing on the Toilet, by performing a mixed-methods case study. We first quantitatively evaluate how effective this technique is by applying statistical causal inference over six years of data about tools used by thousands of developers. We then qualitatively contextualize these results by interviewing and surveying 382 developers, from authors to editors to readers. We found that the technique was generally effective at increasing software development tool use, although the increase varied depending on factors such as the breadth of applicability of the tool, the extent to which the tool has reached saturation, and the memorability of the tool name.**

I. Introduction

Tools can help increase developer productivity by increasing velocity and code quality. For instance, tools can find concurrency bugs [28], reduce the effort to analyze

Two Case Studies of Open Source Software Development: Apache and Mozilla

AUDRIS MOCKUS
Avaya Labs Research
ROY T FIELDING
Day Software
and
JAMES D HERBSLEB
Carnegie Mellon University


Testing on the

Are you tired of
annoyed by figh
Consistent form
what the code do
of personal style
good task for hu
Clang-format pre

```
$ cat fil
int a;//
int bbb;
#define
LOG(INFO)<<".. align operators\n"<<".. and many more things";
$ clang-format file.cc -style Google
int a;     // clang-format can ..
int bbb;   // .. align trailing comments.
#define UNDERSTAND_MULTILINE_MACROS                    \
  int cc;                                              \
  int d;
LOG(INFO) << ".. align operators\n"
          << ".. and many more things";
```

Conveniently integrating with your editor, you can format the current statement or a selected region (available for vim, emacs and eclipse - git/clang-format). You can also reformat unified diffs, e.g. in a CitC client, by:

```
$ g4 diff -du0 | /usr/lib/clang-format/clang-format-diff.py
```

In addition to making the editor-based code development faster and more fun, consistently using clang-

Myths about Case Study Research

- General, theoretical (context-independent) knowledge is more valuable than concrete, practical (context-dependent) knowledge.
- One cannot generalize on the basis of an individual case; therefore, the case study cannot contribute to scientific development.

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

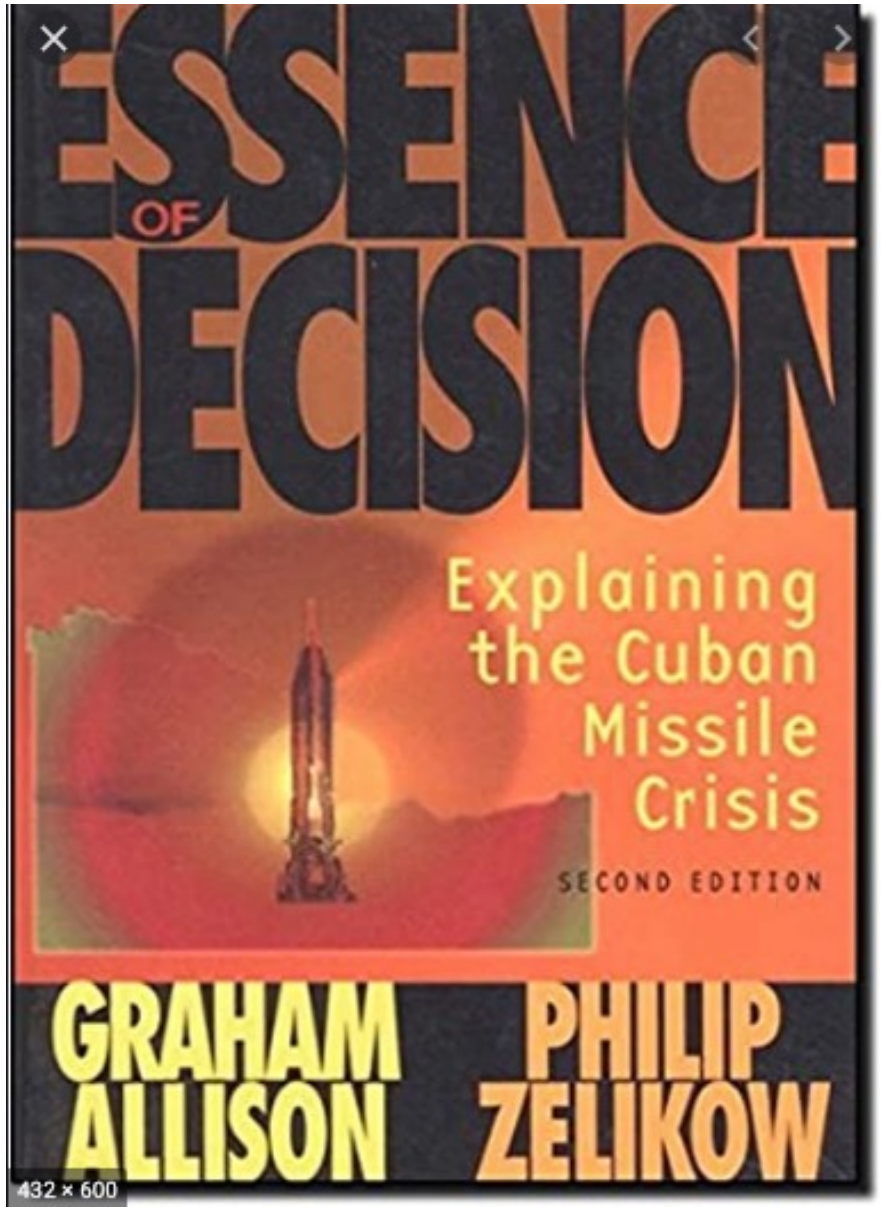
Misuses of the term "Case Study"

- 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

Myths about Case Study Research

- General, theoretical (context-independent) knowledge is more valuable than concrete, practical (context-dependent) knowledge.
- One cannot generalize on the basis of an individual case; therefore, the case study cannot contribute to scientific development.

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



The lessons from the case study are intended to be **generalizable** to foreign affairs Morales and also to a whole variety of complex governmental actions.

Generalization

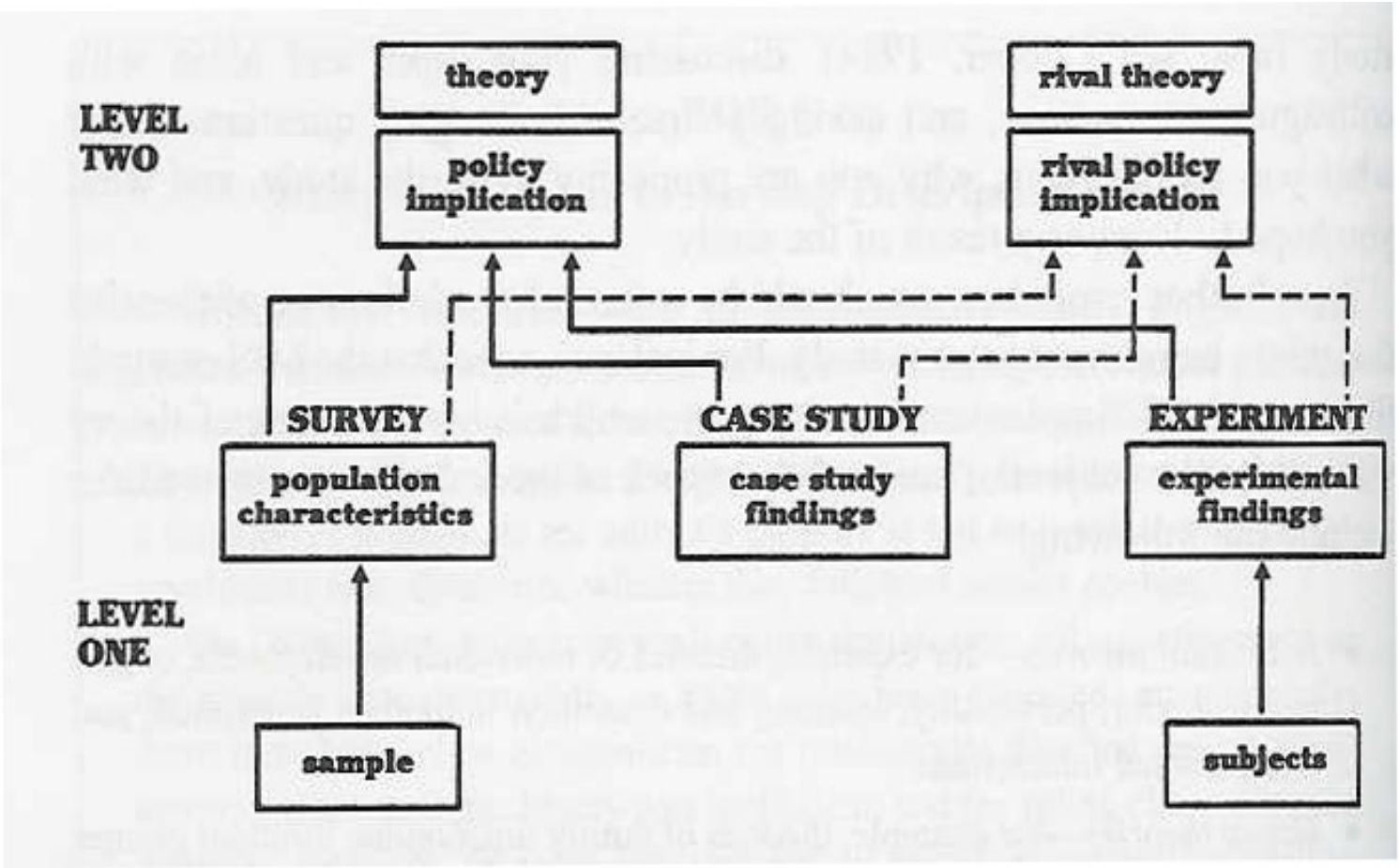
Statistical Generalization

- 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

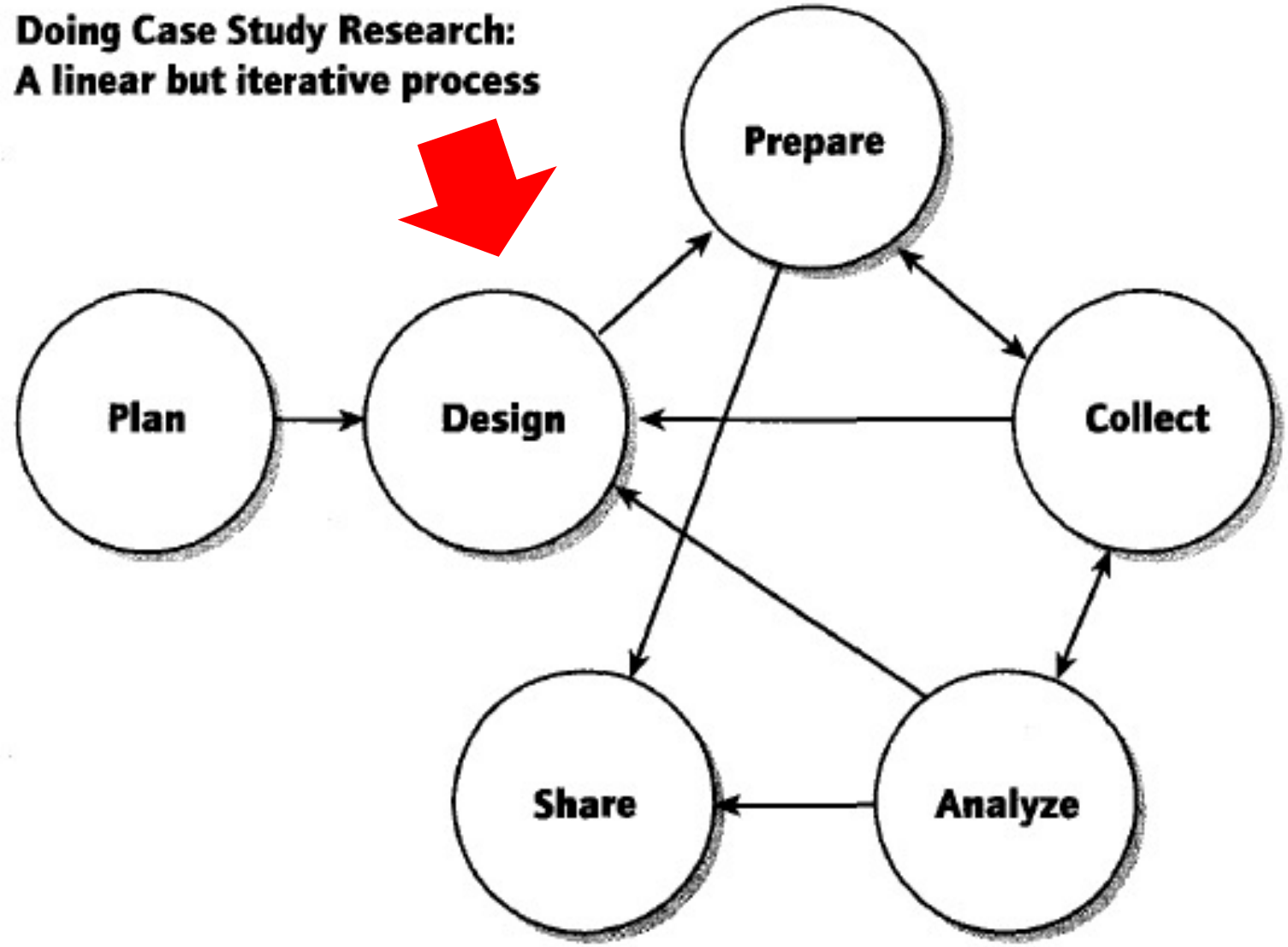
Analytical Generalization

- 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



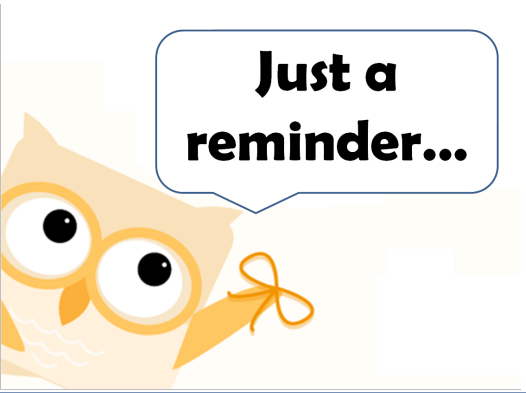
- Is it still a case study when more than one case is included in the same study?
- Do case studies preclude the use of quantitative evidence?
- Can case studies be used to do evaluations?


Doing Case Study Research:
A linear but iterative process
Prepare
Plan
Design
Collect
Share
Analyze

Designing a case study

Planning Checklist

- [x] Pick a topic
- [x] Identify the research question(s)
- [x] Check the literature
- [x] Identify your philosophical stance
- [x] Identify appropriate theories
- [x] Choose the method(s)
- [ ] Design the study
  - [x] Unit of analysis?
  - [x] Target population?
  - [x] Sampling technique?
  - [x] Data collection techniques?
  - [x] Metrics for key variables?
  - [x] Handle confounding factors
- [ ] Critically appraise the design for threats to validity
- [ ] Get IRB approval
  - Informed consent?
  - Benefits outweigh risks?
- [ ] Recruit subjects / field sites
- [ ] Conduct the study
- [ ] Analyze the data
- [ ] Write up the results and publish them
- [ ] Iterate



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 two 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?"

4 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

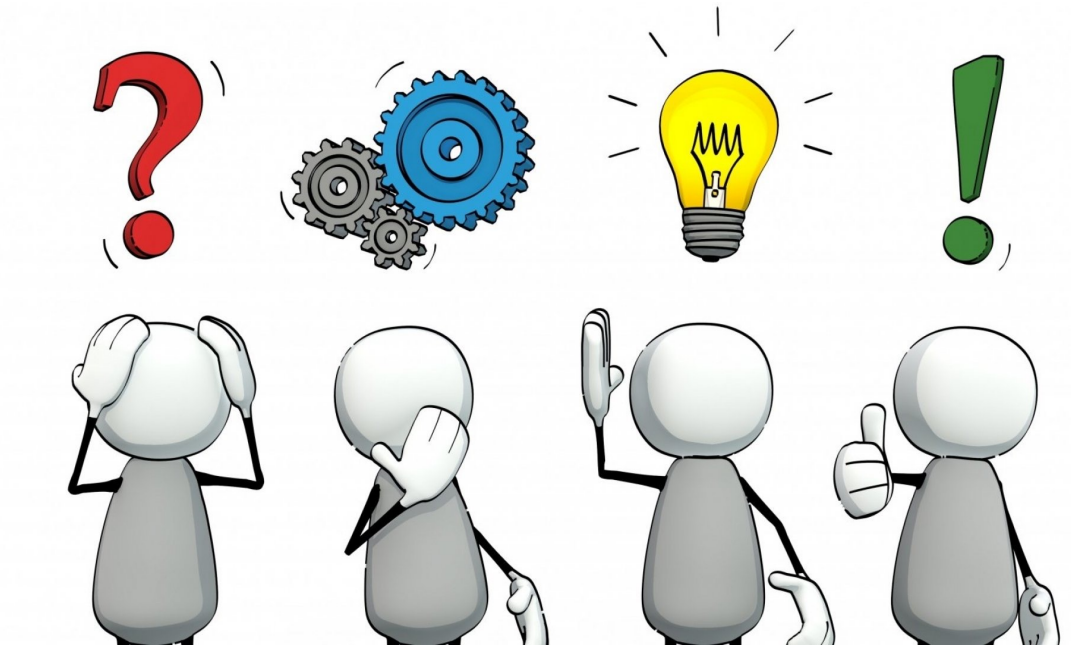
Part 3: Unit of Analysis

- 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 could be…
    - individual developer (focuses on a person's participation in the project)
    - a team (focuses on team activities)
    - a decision (focuses on activities around that decision)
    - a process (e.g. examines how user stories are collected and prioritized)
    - …

Examples of Units of Analysis

- For a study of pair programming

- For a study of software evolution



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

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 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 to generalize to the whole population
  - Qualitative methods cannot use statistical generalization
    - Hence use analytical generalization

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

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

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

Figure 2.3 Case Study Tactics for Four Design Tests

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

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

Figure 2.3 Case Study Tactics for Four Design Tests

Criteria for judging the quality of the research design

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

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

Figure 2.3 Case Study Tactics for Four Design Tests

How can I evaluate a case study?

Same criteria as for other empirical research:

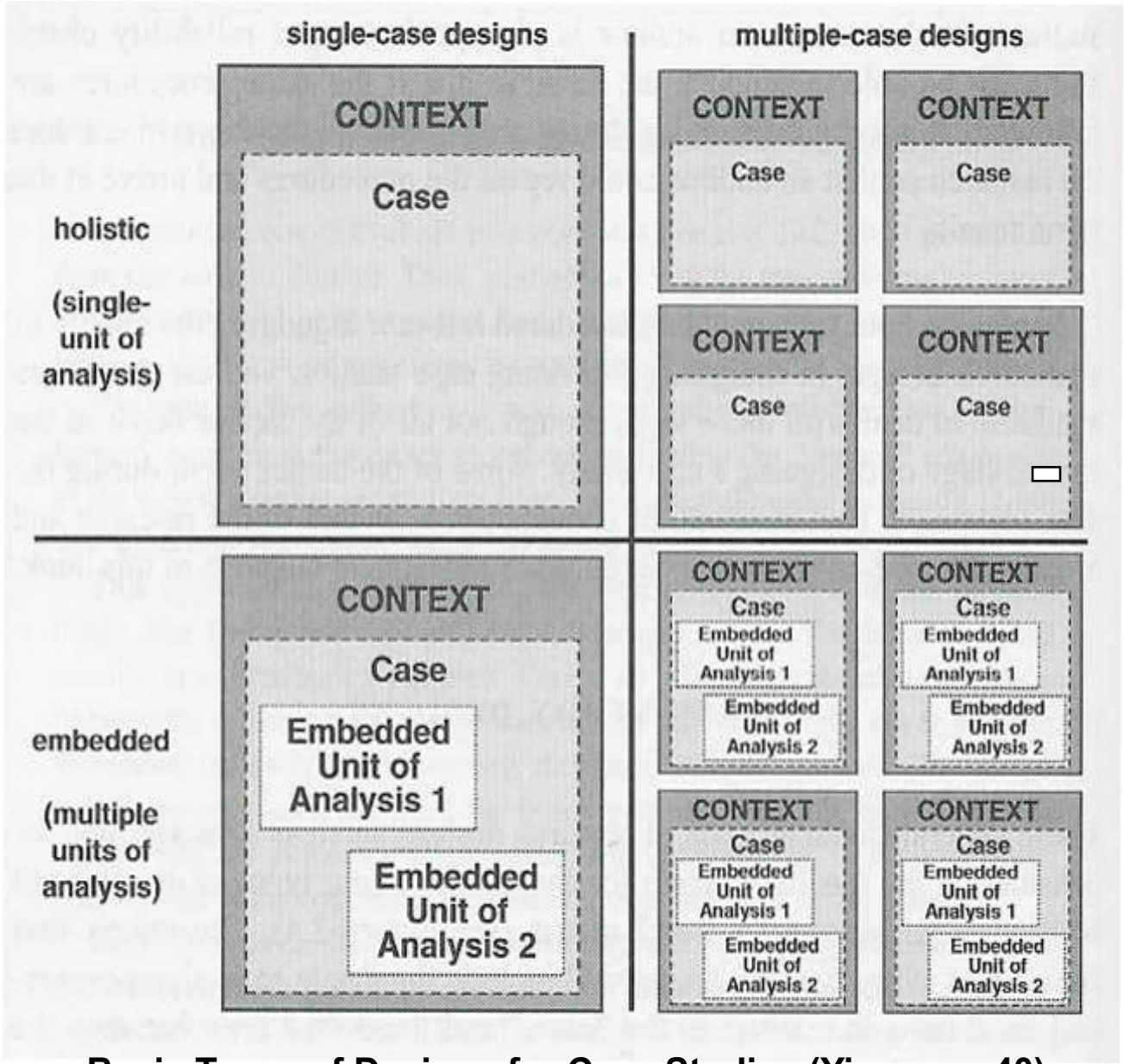
- Construct Validity
  - Concepts being studied are operationalized and measured correctly
- Internal Validity
  - Establish a causal relationship and distinguish spurious relationships
- External Validity
  - Establish the domain to which a study's findings can be generalized
- Empirical Reliability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

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

Figure 2.3 Case Study Tactics for Four Design Tests

Case Study Designs

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



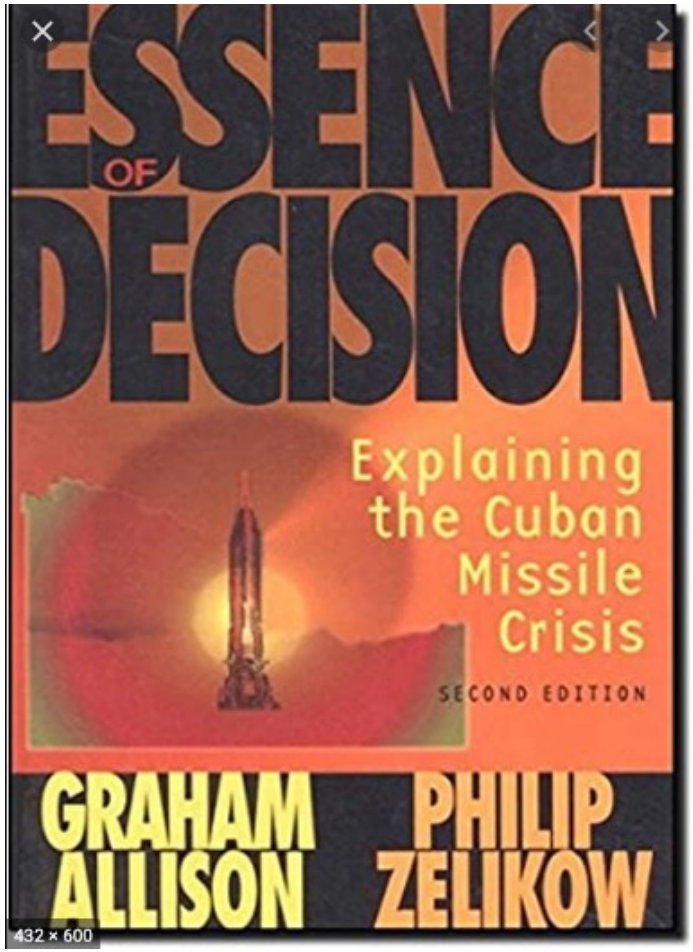
Basic Types of Designs for Case Studies (Yin, page 40)

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

5 reasons to use a single-case design

- It represents the *critical case* in testing a well-formulated theory



5 reasons to use a single-case design

- It represents an extreme or unique case

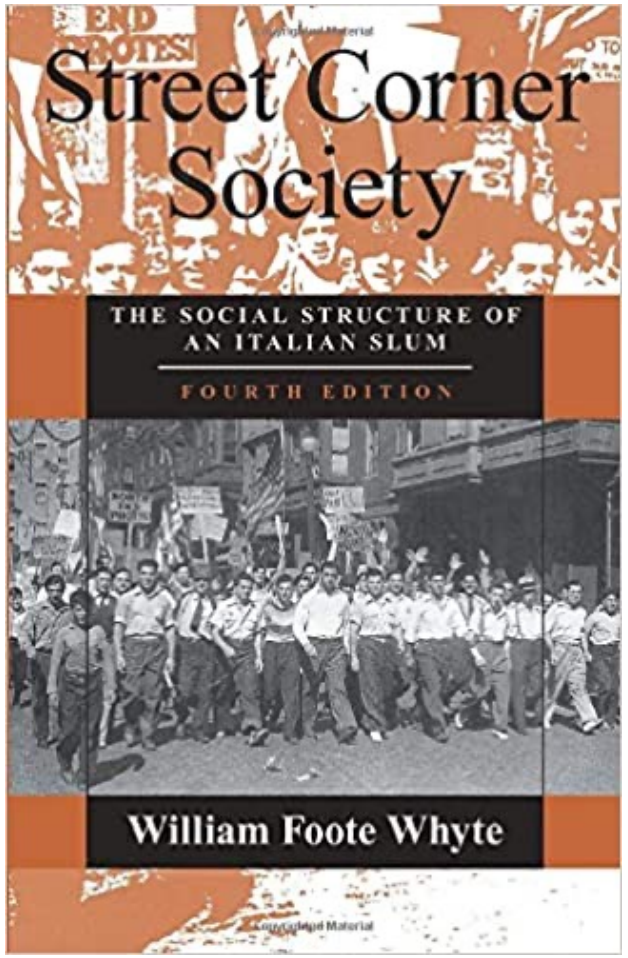
prosopagnosia

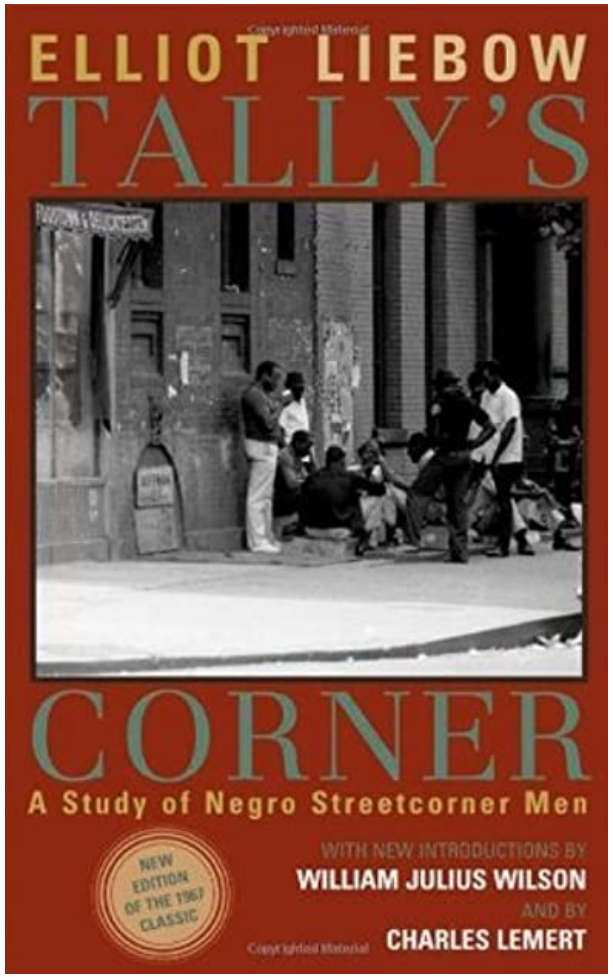
5 reasons to use a single-case design

- It is the *representative* or *typical* case, i.e. informs about common situations/experiences

5 reasons to use a single-case design

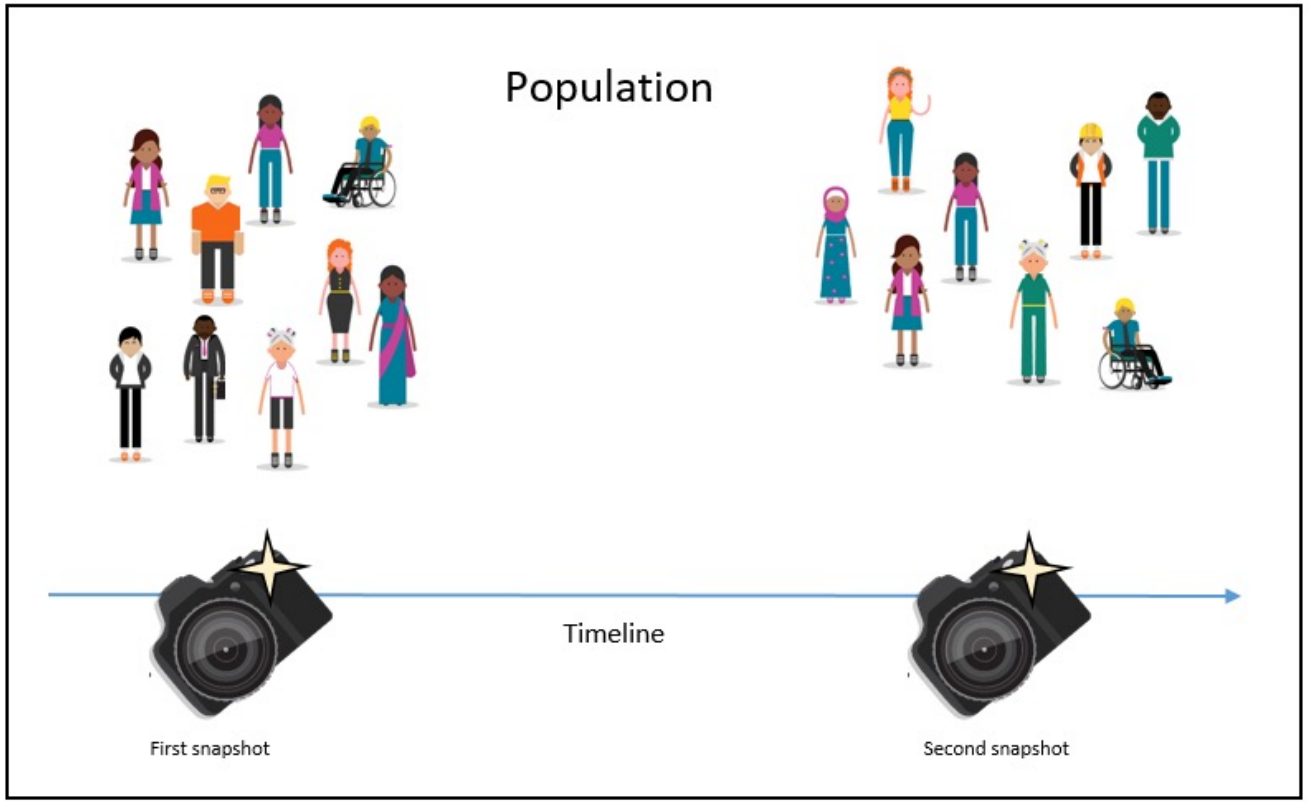
- The case is *revelatory* –a unique opportunity to study something previously inaccessible to observation





5 reasons to use a single-case design

- The case is *longitudinal* – it studies the same case at several points in time



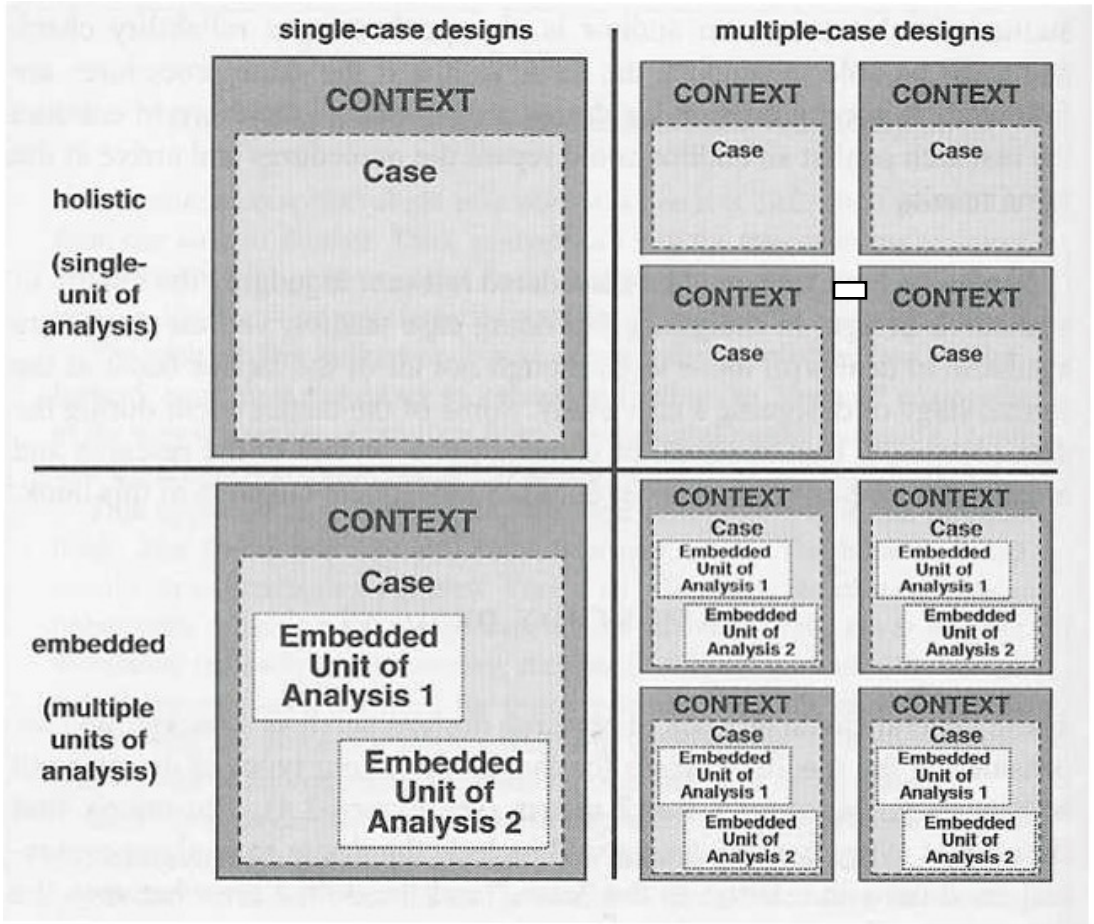
Rationale for Single-Case Designs

- 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

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



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

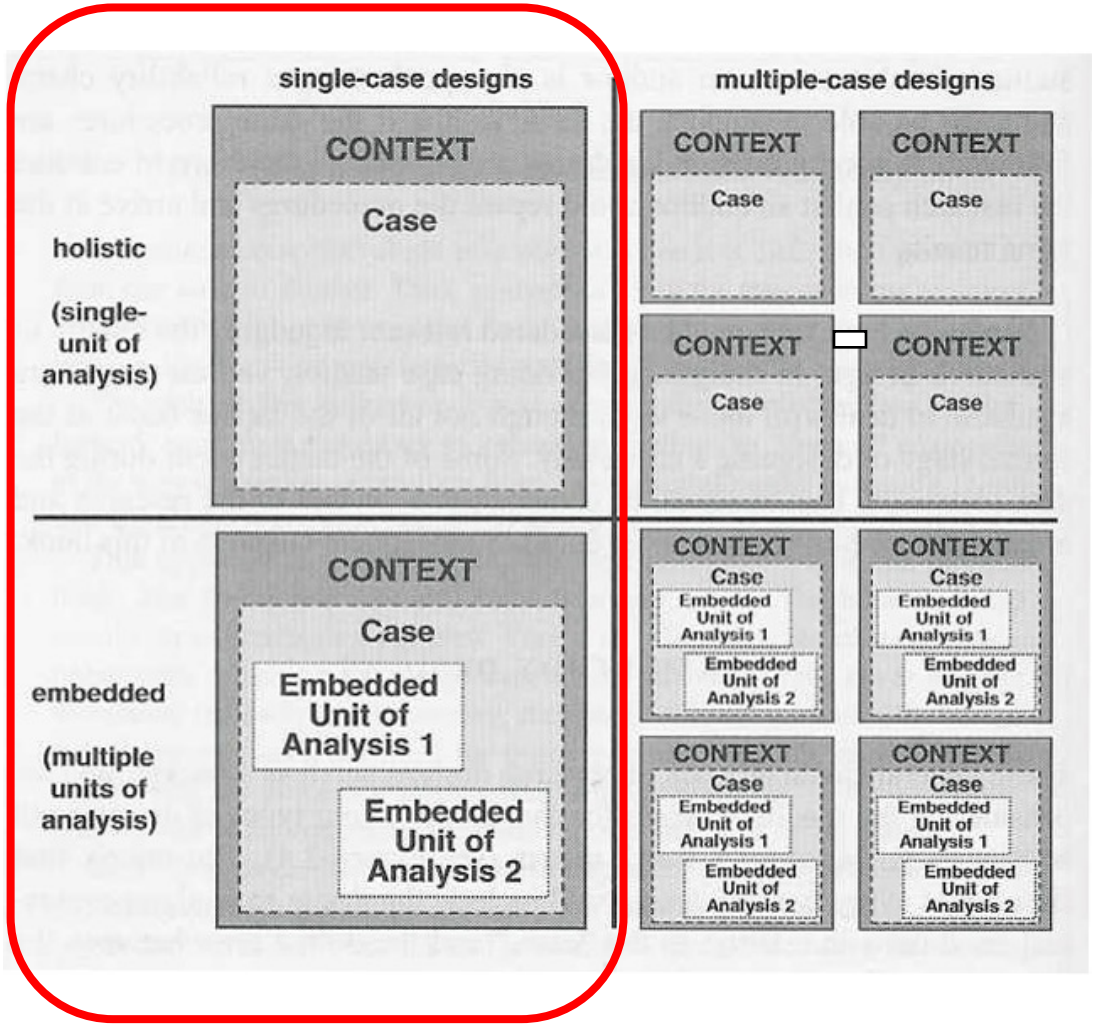
Embedded Designs

- 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

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

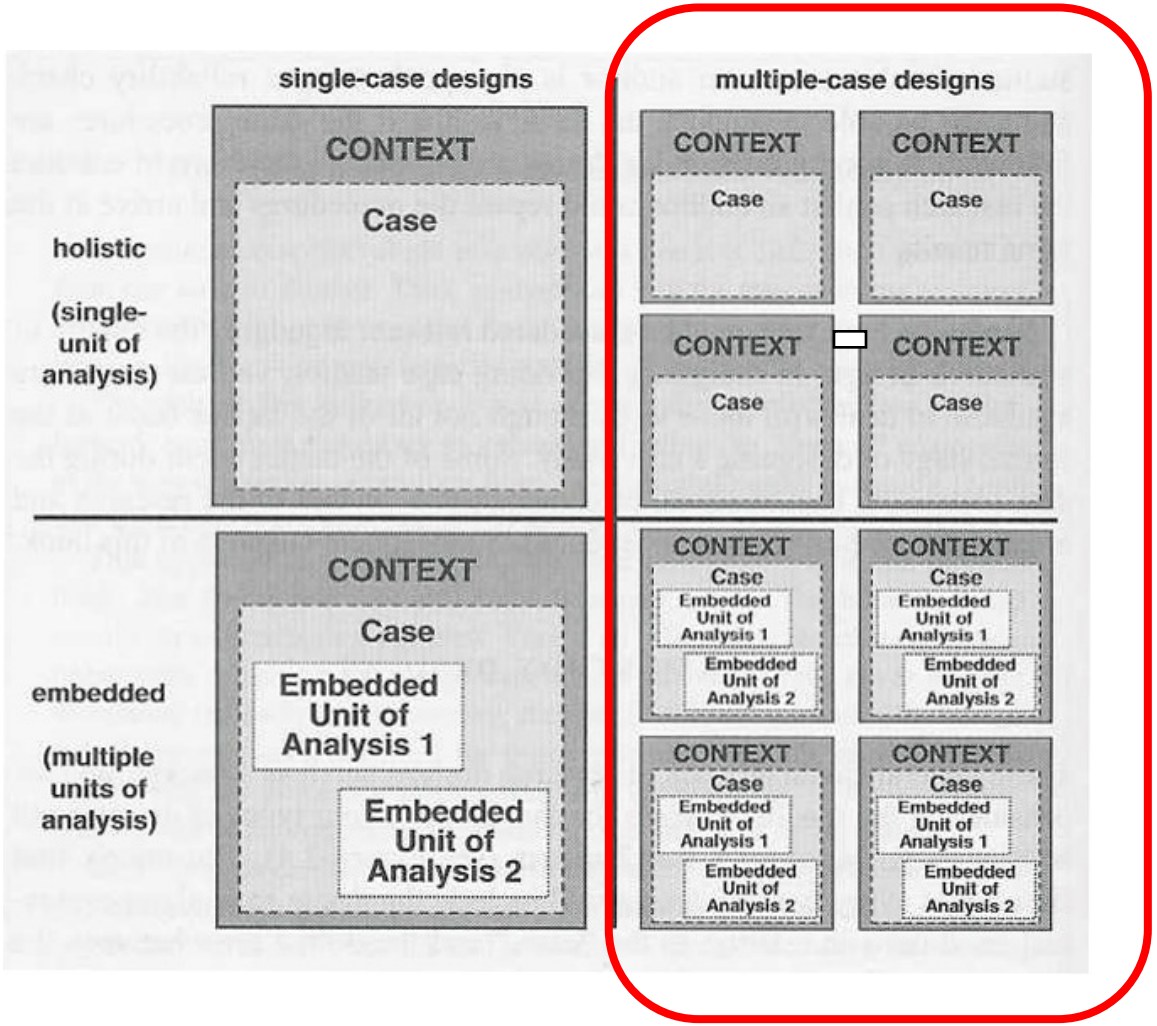
- *Embedded* case study: Involves more than one unit of analysis by paying attention to subunit(s) within the case
  - E.g: a case study about a single organization may have conclusions about the people (subunits) within the organization



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



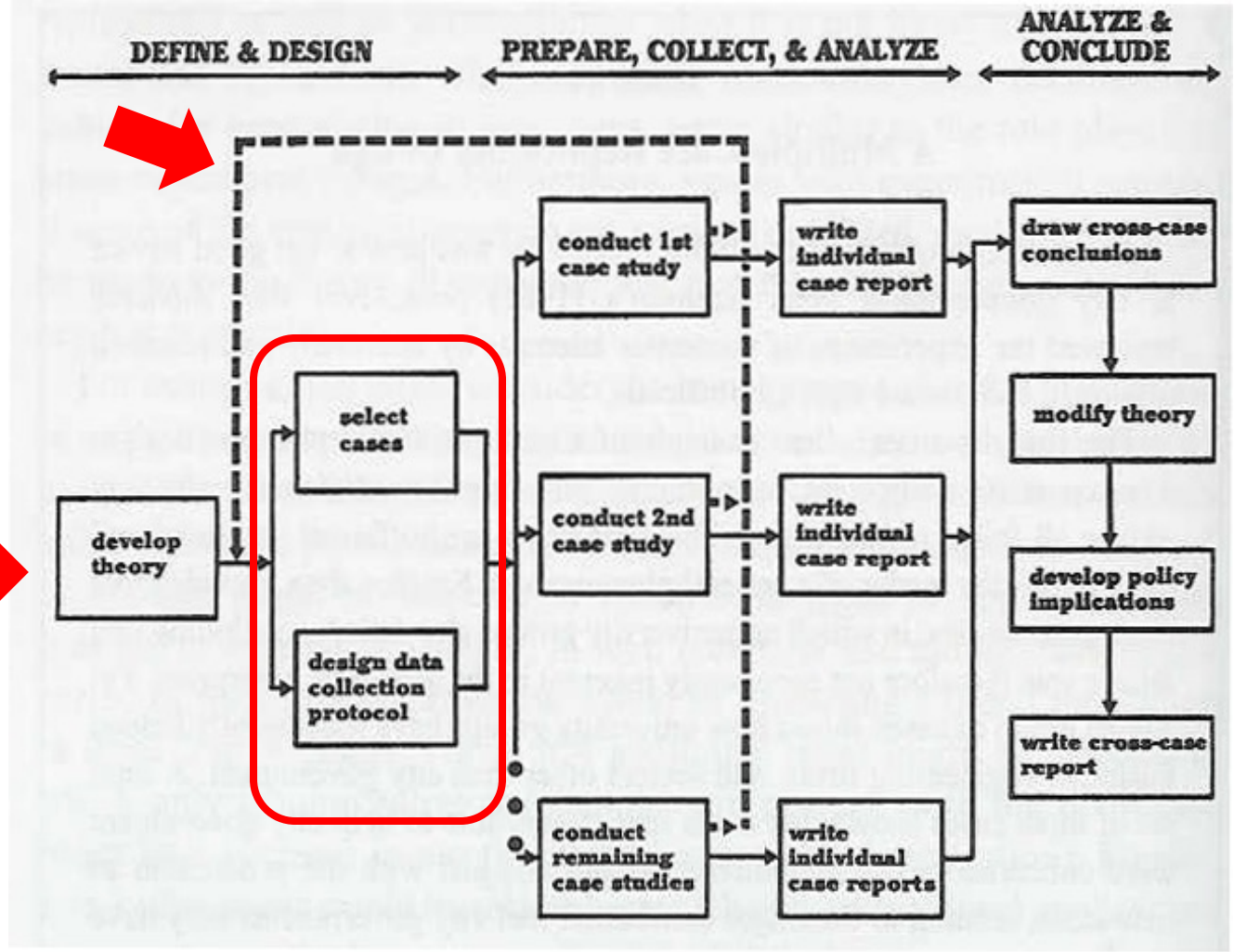
Multiple-Case Designs

- 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

Replication Approach for Multiple-Case Studies

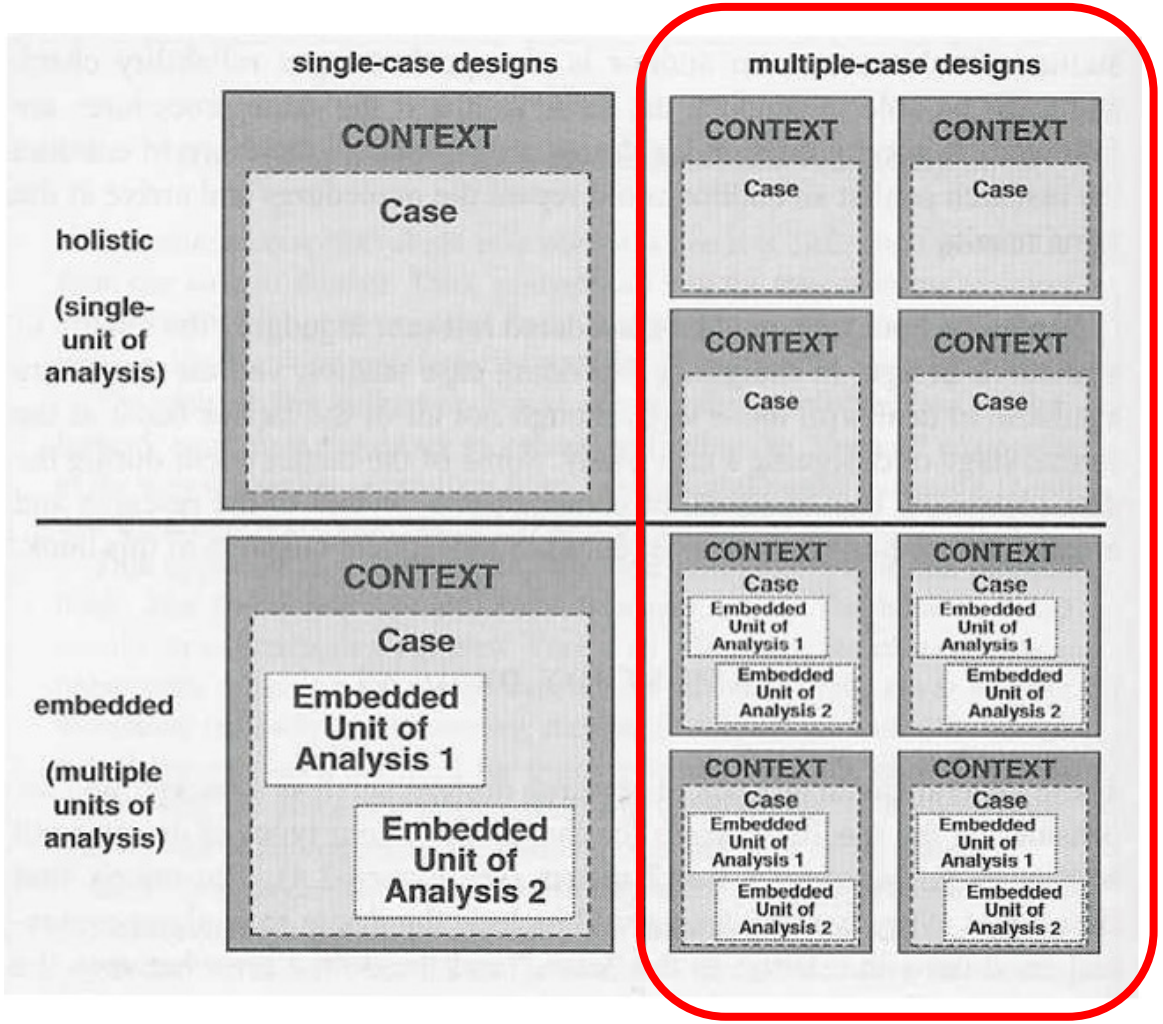


Case Study Method (Yin page 50)

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

Multiple-Case Designs: Holistic or Embedded



- 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

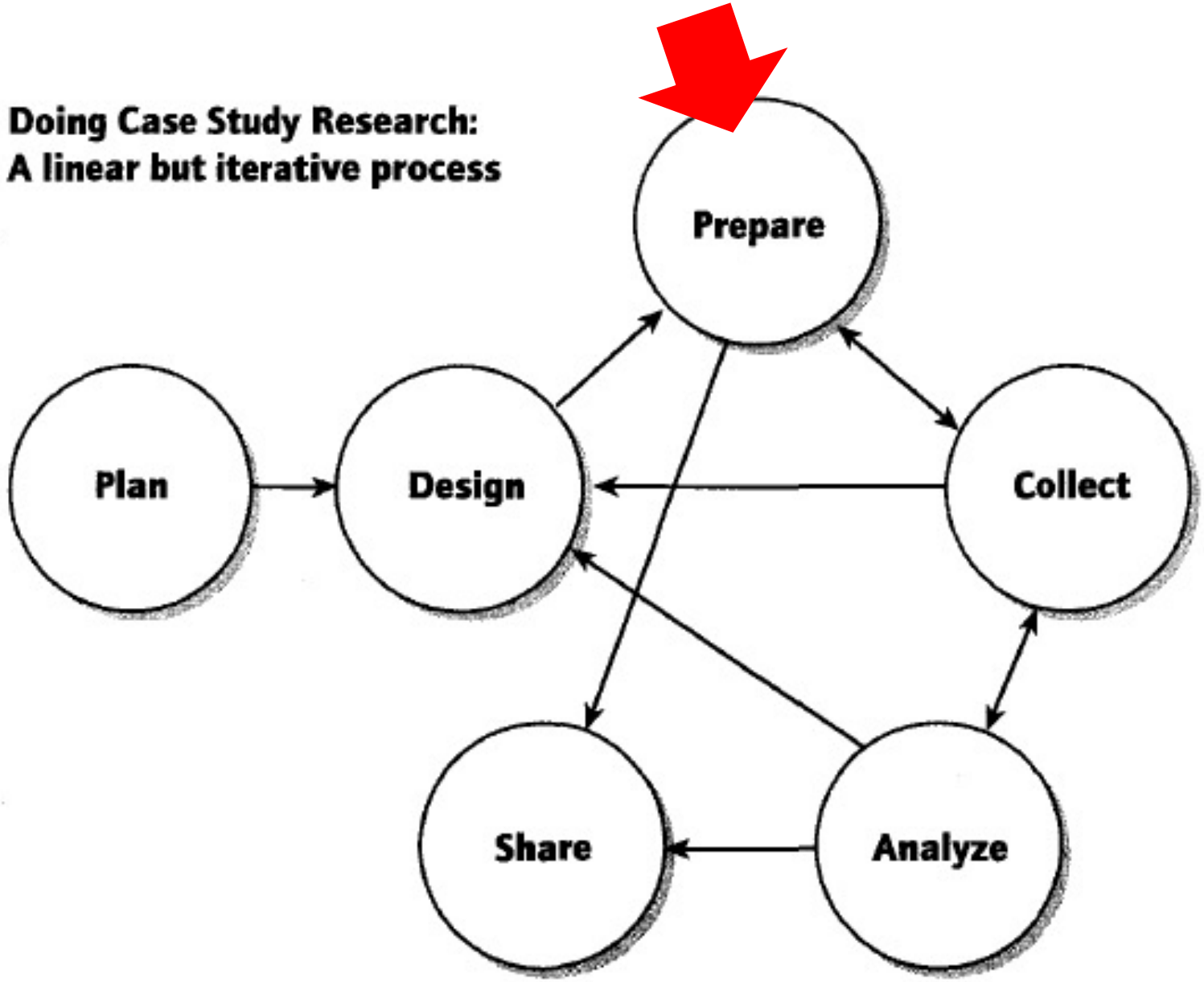
Selecting Case Study Designs – Single or Multiple?

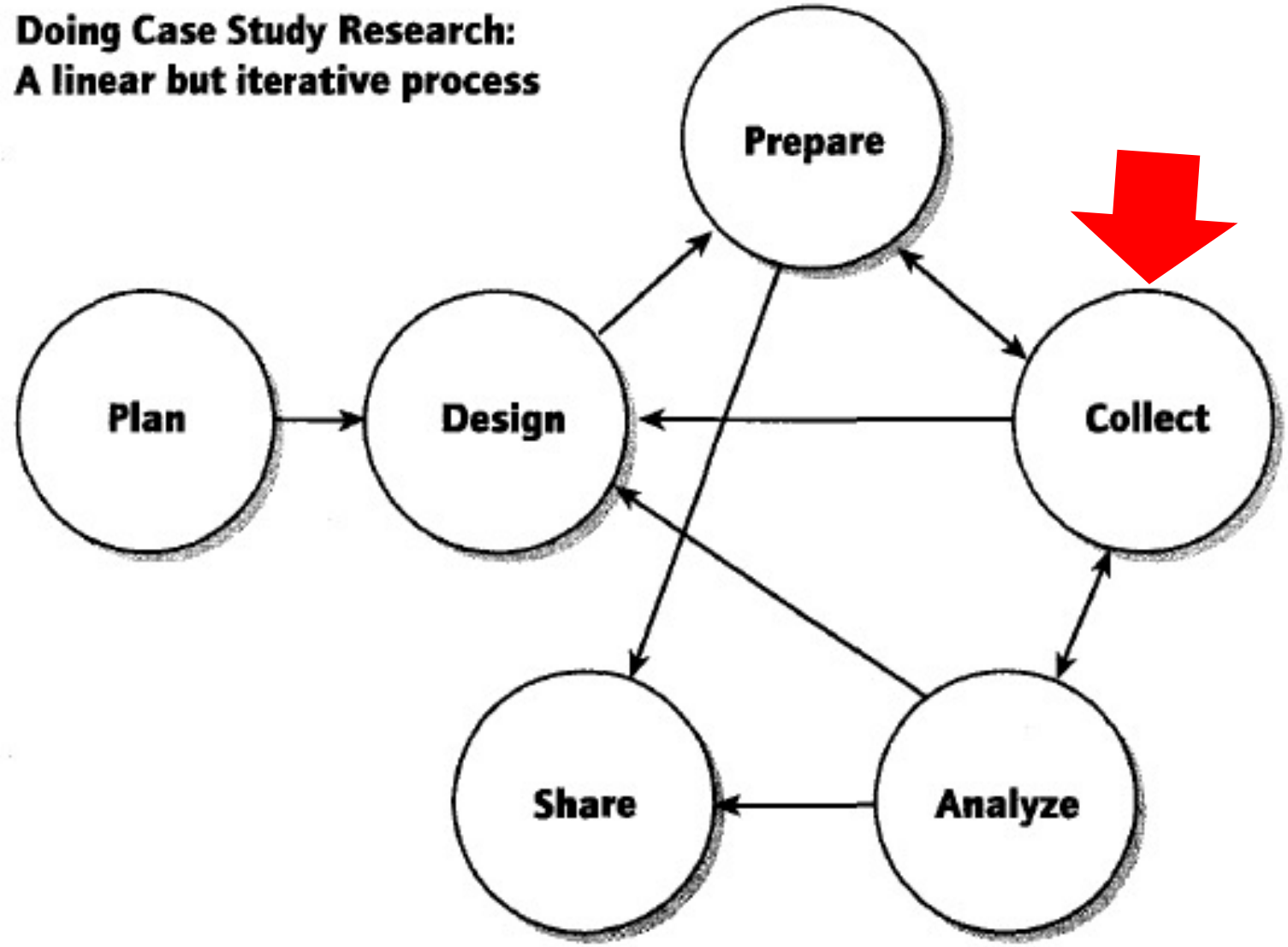
- If you have a choice and enough resources, multiple-case designs are preferred
  - Conclusions independently arising from several cases are more powerful
  - Differences in context of multiple cases with common conclusions improve the generalization of their findings
  - Capability to apply theoretical 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
- In some situations, single-case designs are the best (or only!) choice

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**

Doing Case Study Research:
A linear but iterative process




Doing Case Study Research:
A linear but iterative process
Prepare
Plan
Design
Collect
Share
Analyze

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

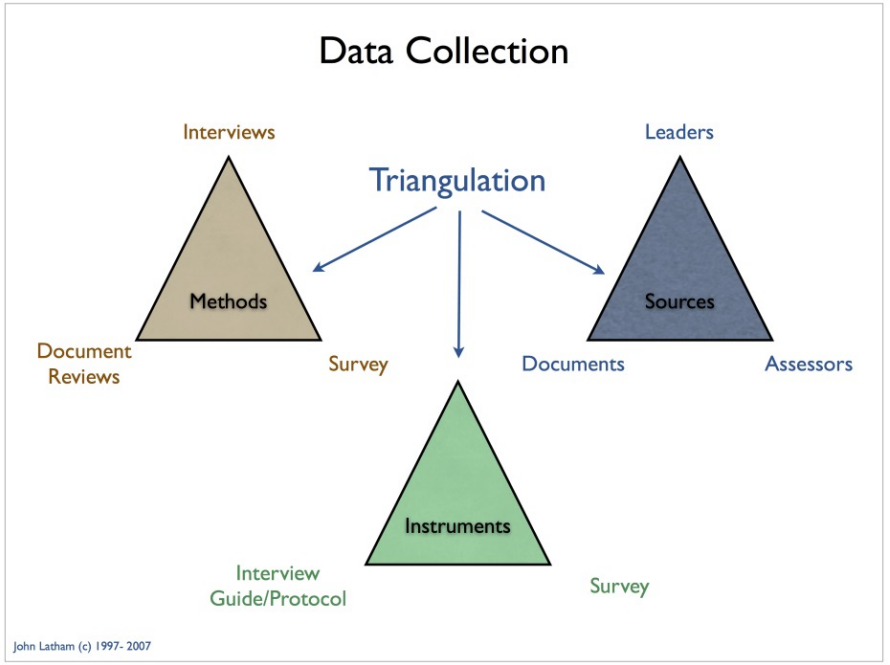
Principles of Data Collection

- 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

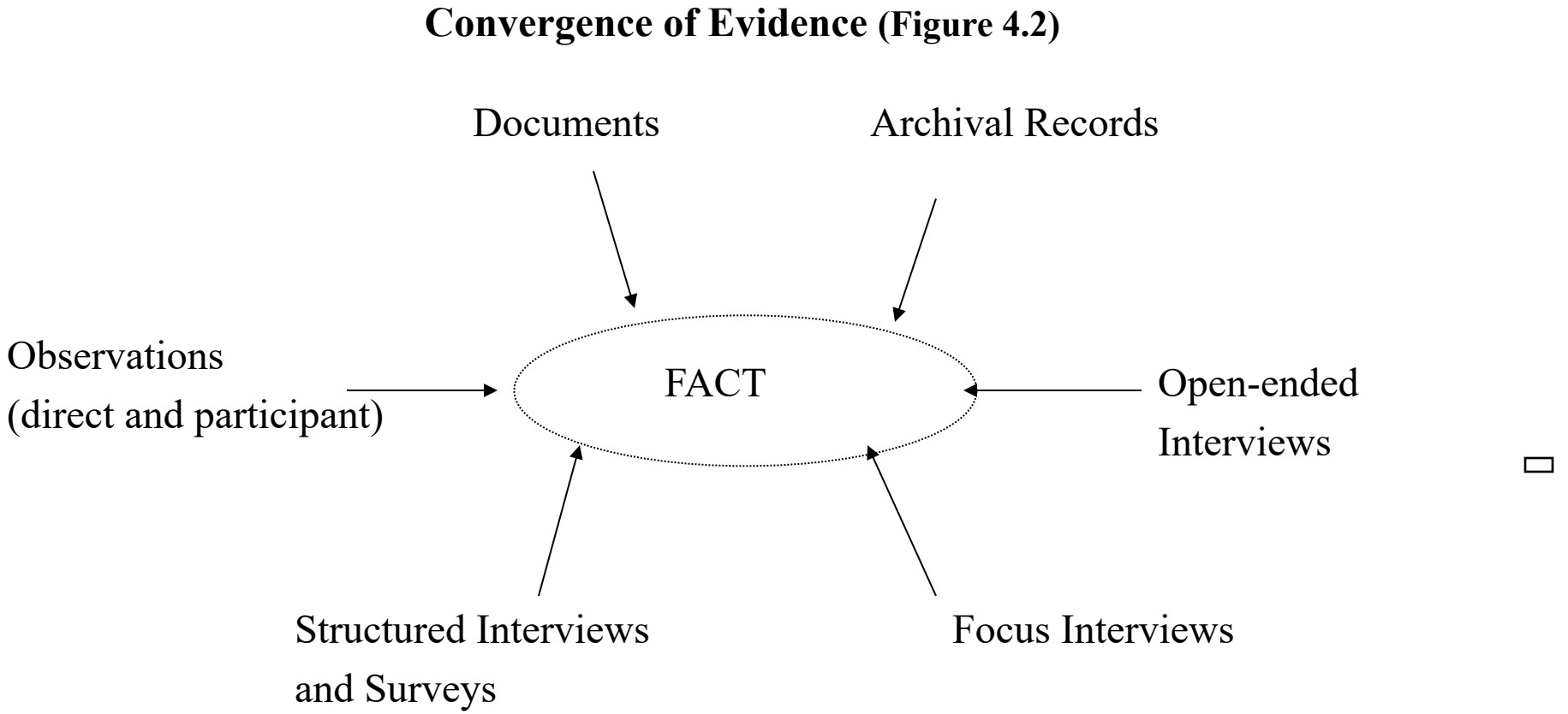
Multiple Sources of Evidence

- 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



Agenda for Today

- Paper reading presentation
- Case studies
- Ethical consideration

Ethical Issues to Anticipate

- Because research involves collecting data from people, Researchers need to:
  - Protect research participants
  - Personal disclosure, authenticity and credibility of research report
  - Develop trust with research participants
  - Promote the integrity of research
  - Guard against misconduct
  - Cope with new problems that emerge

Research Ethics

- Reasons to take ethics seriously:
  - Funding depends on it
  - Relationship with research subjects/organisations depends on it
  - Legal issues (e.g. liability for harm to subjects/organisations)
  - Compliance with privacy and data protection laws
  - …and it's the right thing to do!
- Institutional Review Boards (IRB)
  - Approval usually needed for all studies involving human subjects
  - Every IRB has it's own rules…
    - A study approved at one university may be disallowed at another!
    - Design of the study might have to be altered
  - Institutional research funding may depend on this process!
  - Note: guidelines from other fields may not apply to Software Engineering
    - E.g. use/ownership of source code
    - E.g. effect of process improvement on participants

What is the IRB

Institutional review boards (IRBs)

Also called research ethics committees, IRB's provide protection for human research participants through review of the ethical acceptability of proposals for human research.

What does it consist of?

5+ sufficiently qualified members

Members with diverse experience and expertise to safeguard subjects' rights and welfare and to evaluate research acceptability.

At least one member knowledgeable about any regularly researched vulnerable groups.

Can include specialized experts in relation to particular cases.

According to the NIH

StudyFind

Why do we need the IRB?

StudyFind

1 Functions and Operations

Write procedures for researchers to report issues

Review and amend study procedures

Approval of procedures by majority vote

2 Review of studies

Require informed consent and documentation

Approve, Modify, Disapprove Research

Review research annually

3 Authority

IRB must approve research to move forward.
IRB can suspend or terminate research for serious harm or noncompliance.

4

Keep records of research proposals, meetings ,etc.

According to the NIH


UNIVERSITY OF TORONTO
Administrative Web Services
To protect the security of your data, this application will automatically terminate in 03:59:33. Unsaved work
Welcome
Inbox
My HR Self-Service
My Research
Expense Reimbursement
RAISE Help Page
Inbox
Applications & Agreements
My Research On Line
Human Ethics Protocols
My Research > Human Ethics Protocols > PI > My Human Research Protocols

Informed Consent

- Elements
  - Disclosure - participants have full information about purpose, risks, benefits
  - Comprehension - jargon-free explanation, so participants can understand
  - Competence - participants must be able to make rational informed choice
  - Voluntariness - no coercion or undue influence to participate
  - Consent - usually indicated by signing a form
  - Right to withdraw
    - participant can withdraw from study at any point without having to give reasons
    - Participants can request their data to be excluded (might not be possible!)
- Challenges:
  - Student participants
    - Perception that their grade will be affected if they don't participate
    - Perception that it will please the course instructor if they participate
  - Industrial participants
    - Perception that the boss/company wants them to participate

Ethical Issues: Beginning the Study

- In the Research Problem
  - Identify a problem that will benefit individuals being studied
- In the Purpose and Questions

Convey the purpose and sponsors of the research to participants

- Do not pressure participants into signing consent forms, obtain informed consent from participants
- Respect norms and charters of indigenous cultures

Ethical Issues: Collecting the Data

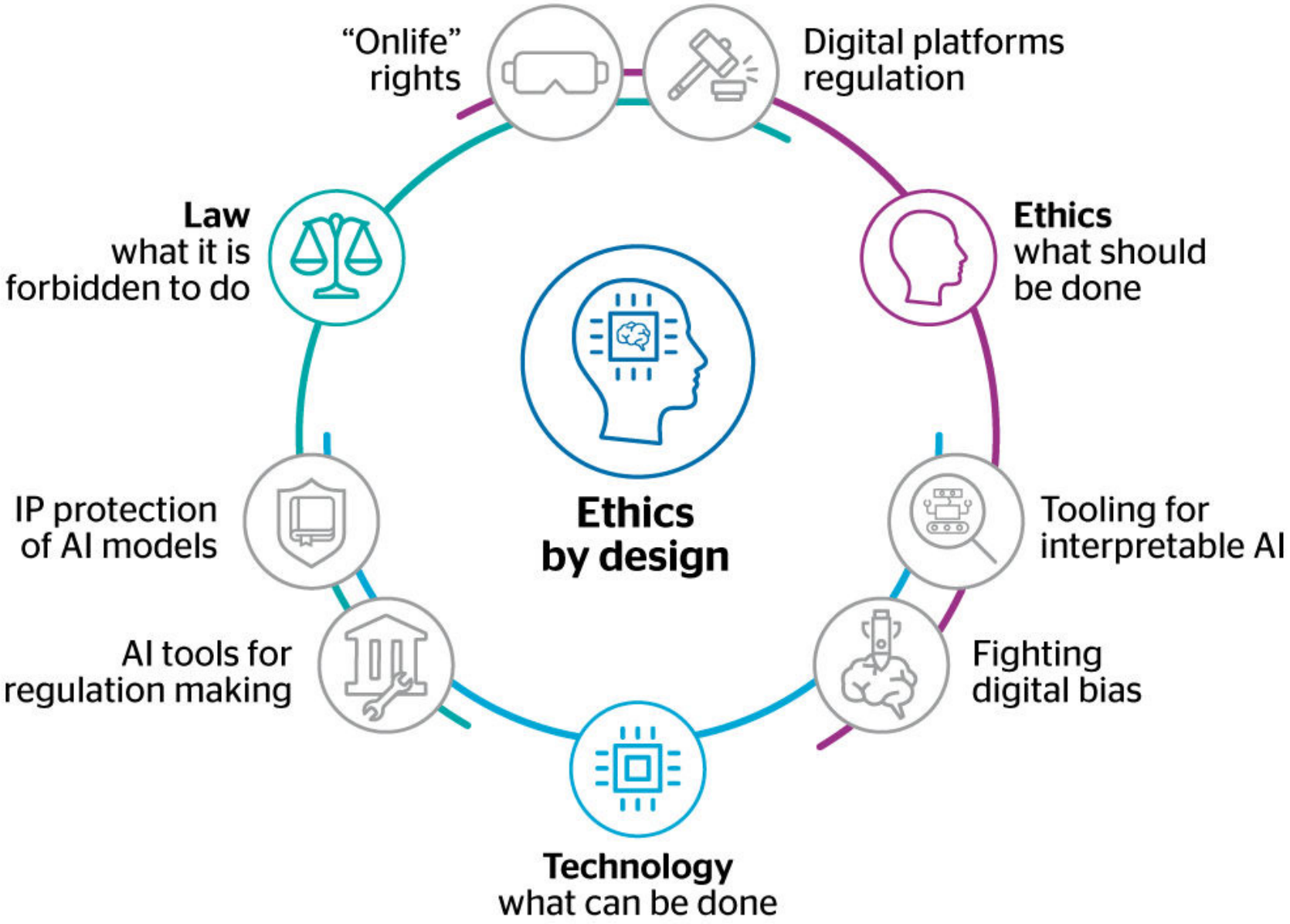
- Respect the site, and disrupt as little as possible
- Make sure all participants receive the benefits
- Avoid deceiving participants
- Respect potential power imbalances, consider reciprocity
- Avoid exploitation of participants
- Avoid collecting harmful information, do not put participants at risk
- Respect vulnerable populations
- Address issues of confidentiality
- Interview with sensitivity
- Anticipate issues that may arise

Ethical Issues: Data Analysis

- Avoid going native, do not take sides or disregard data that proves or disproves personal hypotheses
- Avoid disclosing only positive results, data analysis should reflect the statistical tests and not be underreported
- Respect the privacy of participants:
  - Protecting anonymity of participants
  - Storing data and destroying it after a set time
  - Planning for ownership of the data and not sharing data with others

Ethical Issues: Reporting, Sharing, and Storing Data

- Do not falsify authorship, evidence, data, findings or conclusions
- Do not plagiarize
- Avoid disclosing information that would harm participants
- Communicate in clear straightforward, appropriate language
- Share data with others (example: stakeholders, participants)
- Keep raw data and other materials for a reasonable period of time
- Do not duplicate or piecemeal publications
- Complete proof of compliance with ethical issues and lack of conflict of interest
- Understand who owns the data


"Onlife" rights
Digital platforms regulation
Law what it is forbidden to do
Ethics what should be done
Ethics by design
IP protection of AI models
Tooling for interpretable AI
AI tools for regulation making
Fighting digital bias
Technology what can be done

An Ethical Dilemma..

You are doing a study of how junior analysts use new requirements tool at a leading consultancy company. As part of informed consent, staff are informed that they will remain anonymous. During the study, you notice that many of the analysts are making data entry errors when logging time spent with clients. These errors are causing the company to lose revenue. Company policy states clearly that workers salaries will be docked for clear mistakes leading to loss of revenue.

Questions:

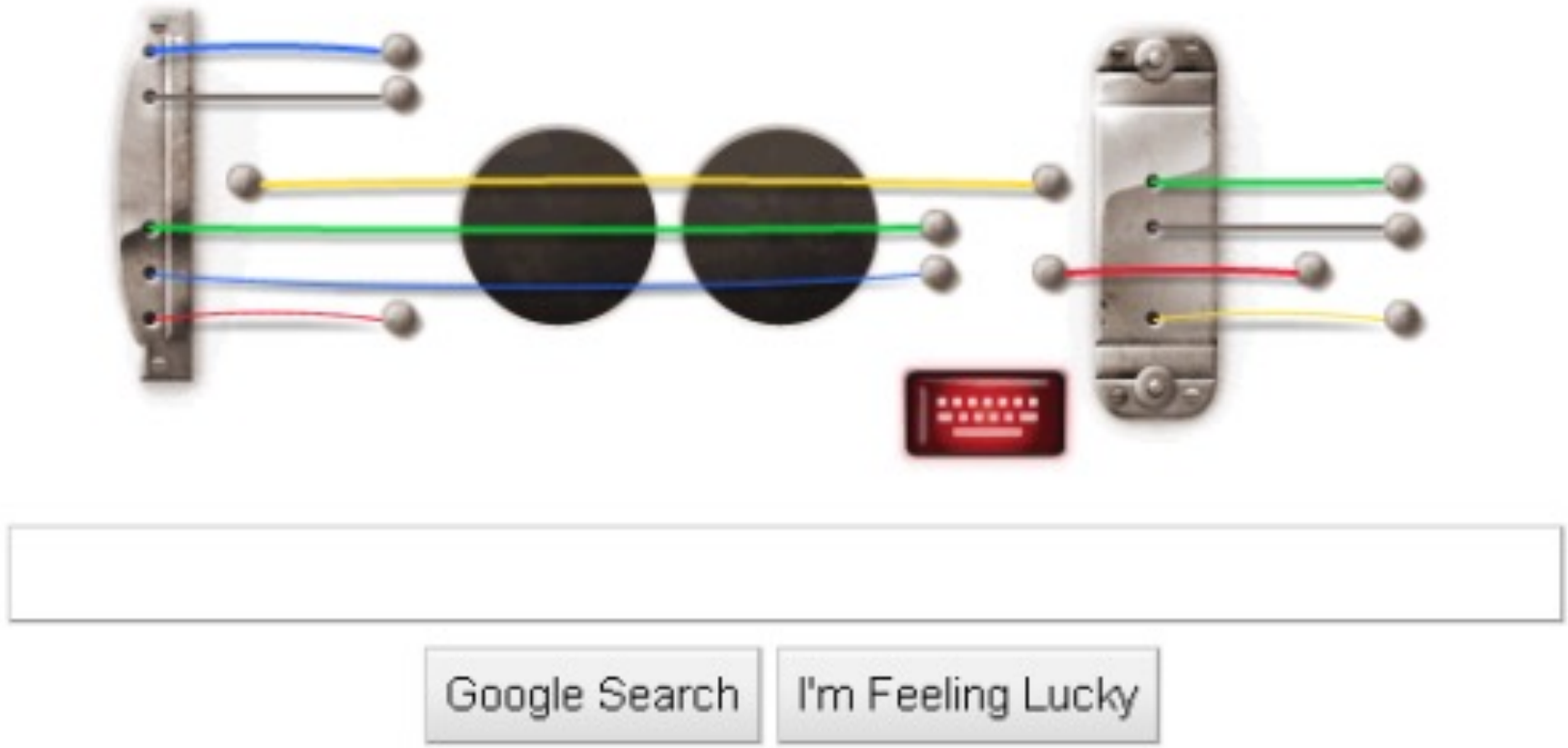
- Would you alter the results of your study to protect the people who helped you in the study?
- How can you report results without causing harm to the participants?
- Would you cancel the study as soon as this conflict of interest is detected?

Confidentiality

- Protecting Anonymity
  - Do not collect any data (e.g names) that allow participants to be identified
  - But you need a signed consent form, so…
  - Sever participants' identity from their data before it is stored and analyzed
  - Researcher-subject interactions should be held in private
- Protecting the data
  - Consent form states who will have access to the data, and for what purpose
    - Do not stray from this!
  - Raw data should be kept in a secure location
  - Reports should only include aggregate data
- Exceptions:
  - When it is impossible to identify individuals from the raw data
  - When more harm results from maintaining confidentiality than breaching it

Why Software Engineering Courses Should Include Ethics Coverage?

- Software helps shape, not just reflect, our societal values
- Examples:
  - How many cars or rockets are made today that do *not* depend upon critical software for their safe operation?
  - How many bridges are built today without the use of sophisticated computer programs to calculate expected load, geophysical strain, material strength and design resilience?



"Update Jun 17: Wow—in just 48 hours in the U.S., you recorded 5.1 years worth of music—40 million songs—using our doodle guitar.
And those songs were played back 870,000 times!"

https://blog.rescuetime.com/google-doodle-strikes-again/

Les Paul Doodle

- Likely designed in days, side project
- Used by users for 5.3 million hours (8 lifetimes)
- Questions: Time sink, lost productivity? Negative or positive net contributions to the world? Who should consider cost/benefits? Based on what principles?

EA calls its loot boxes ‘surprise mechanics,’ says they’re used ethically

80

‘People like surprises,’ executive tells UK Parliament

By Ana Diaz | @AnaLikesPikachu | Jun 21, 2019, 9:10am EDT

SHARE

The Morality Of A/B Testing

Josh Constine @joshconstine / 8:50 PM PDT • June 29, 2014

Facebook COO Sheryl Sandberg apologizes for psychological News Feed experiment

by NICK SUMMERS — Jul 2, 2014 in FACEBOOK

The Ethics of Software Development with Uncle Bob Martin

October 19th 2020 2

;

- https://hackernoon.com/the-ethics-of-software-development-with-uncle-bob-martin-6f153t2r