Building Theory With Case Studies: notes for SE&D Research

Zoe Szajnfarber, Associate Professor

Engineering Management & Systems Engineering and International Affairs
George Washington University

This lecture builds on:

1. Material that is part of the GWU class EMSE 8000
2. A qualitative methods workshop conducted at CESUN 2016 (joint with E. Gralla)
3. The paper: “Qualitative Methods for Engineering Systems: Why we need them and how to use them” (in review) (joint with E. Gralla) – provided read-ahead

Disambiguation

Case studies as an empirical basis for building (and/or elaborating) your theory

vs.

A ***case study*** used to prove that your new method works as advertised.

Agenda

- When should you use case study methods?
- Where to start: Framing a question vs. testing a hypothesis
- Qualitative sampling: how do you pick cases/population?
  - Levels of selection and how to count "N"
  - N != N, and depth vs. breadth
  - Quasi-experimental design vs. replication logic
  - Statistical vs. Analytic Generalizability
- Scoping and conducting data collection
- Analysis strategies for inductive inference
  - Overview of process
  - Where the magic happens and how to be sure leaps are valid
- How to judge if the output of a case study is "good"?

When to use case study methods

A spectrum of theory building options

Observe the world

Qualitative
(Case studies, ethnographies etc.

1. Pick cases
2. In-depth observation population/artifacts
3. Infer abstracted patterns
4. Output: (tentative) Explanations

Quantitative
(Surveys, econometrics, big data etc)

1. Pick cases/population
2. Operationalize abstracted measures
3. Measure effects (quantitative tools: networks, regression etc)
4. Output: clean measures of correlation; argument for causation

Manipulate (experiment)

Experiments
(Field, lab, serious games etc)

1. Represent phenomenon
2. Select actors (subjects)
3. Pick treatment/controls set of runs etc.
4. Measure effects
5. Output: clean measures of causation (here)

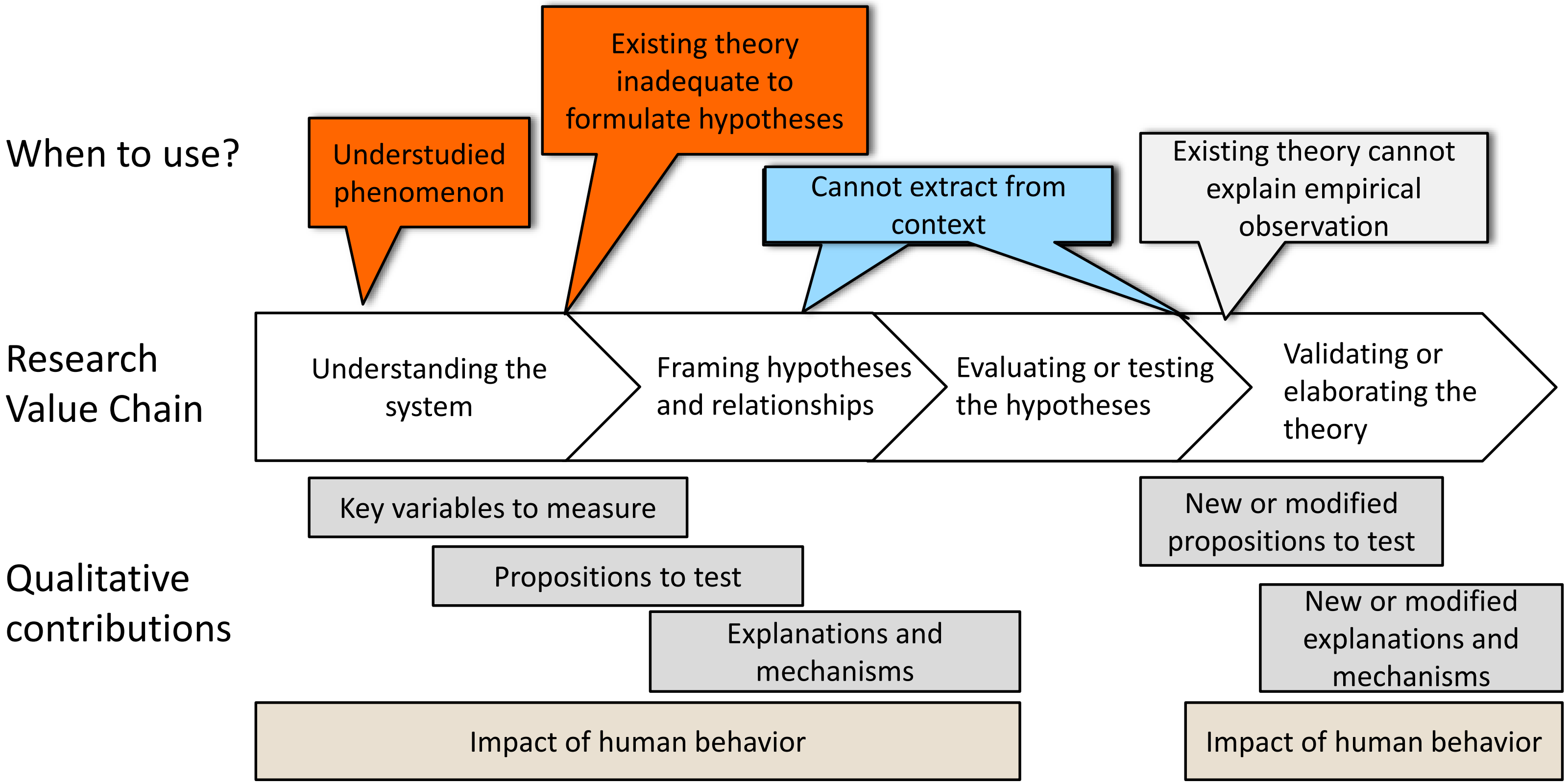
Models
(Formal, simulation etc.)

1. Represent phenomenon, actors (subjects) and set of interventions
2. Pick set of runs to compare
3. Measure effects (quant tools)
4. Output: complex measures of relationships (can get causation)

Each of strengths and weaknesses and an important role to play in studying and understanding the design and designers (and the world)

Where (depth) case studies help most

(hint: not everywhere)



Where to start: Framing a question vs. testing a hypothesis

The Hypothesis Trap: Questions are OK

- Engineers are often taught that objective research is framed around clear and testable hypotheses.
- However, in nascent, nebulous research areas, where case studies are most helpful, a focus on hypotheses can be harmful:
  - They can limit what you observe… and you might miss critical/valuable insights.
  - Can lead to confirmation bias, or frequent null results
- It is ok (and preferable) to start with a broad question and refine it based on what you see.
  - NB: this makes the <u>design</u> of the research critical to validity!!

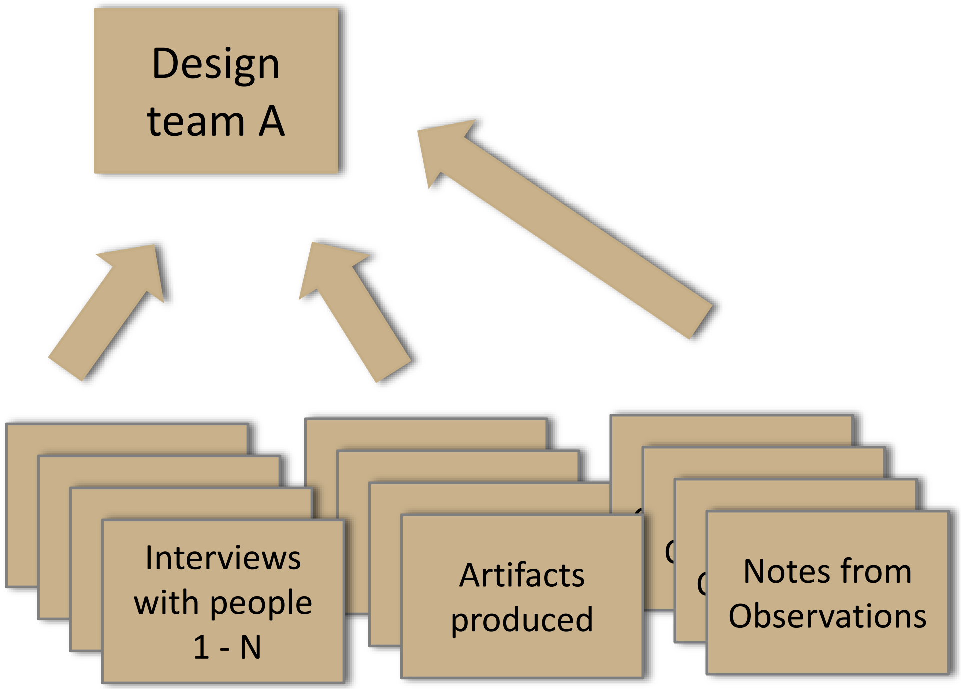
Qualitative sampling: how do you pick cases/population?

where (much of) validity comes from

Defining “selecting cases:” N confusion

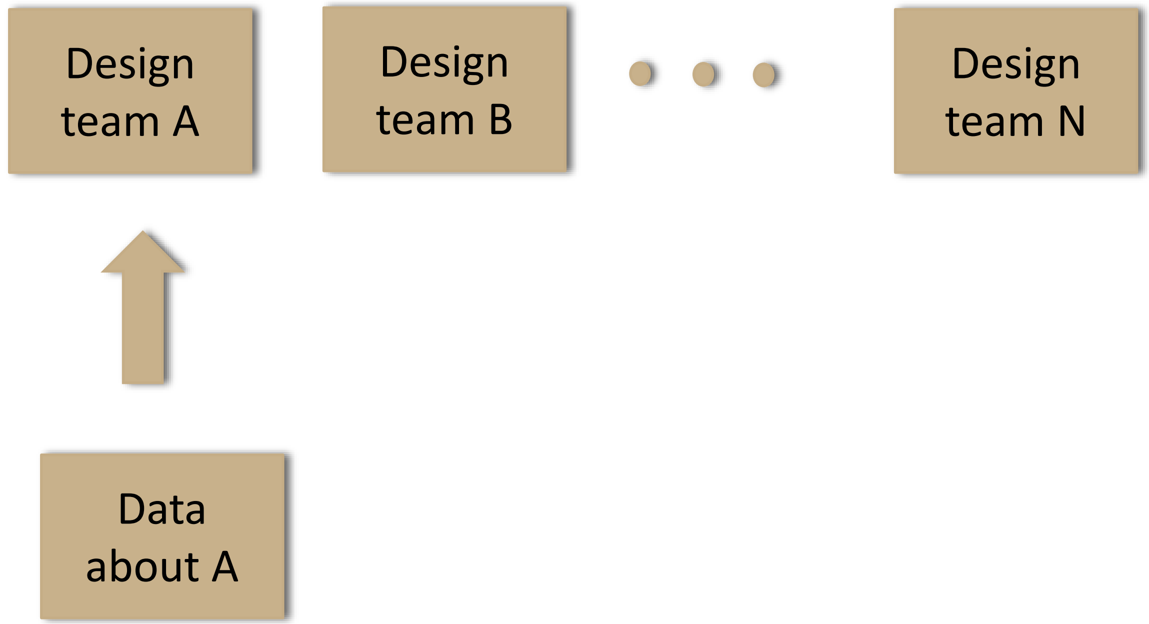
Points of comparison

Sources of data about each “point”



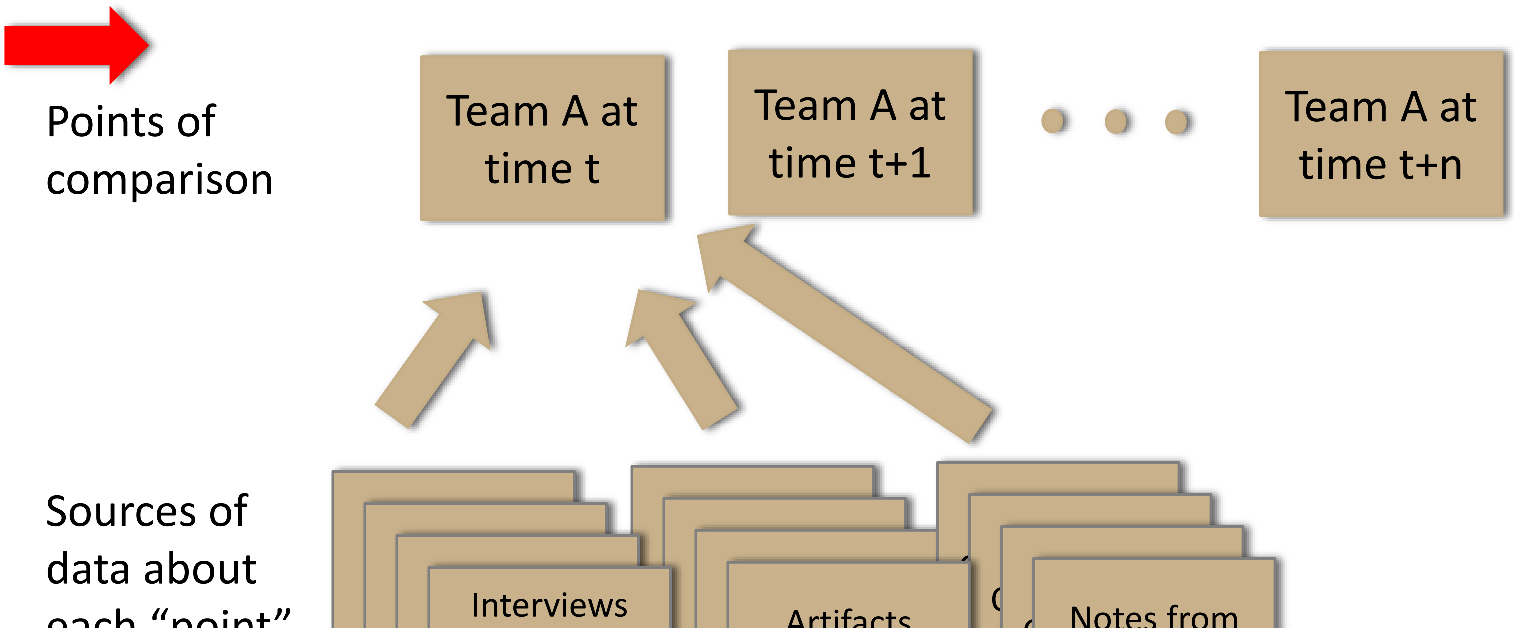
Defining “selecting cases:” N confusion

Points of comparison



Sources of data about each “point”

Defining "selecting cases:" N confusion



How many "N"?
Does it matter if all the team's are in the same organization?
If I study 1 team over 3 periods, is that the same as 3 teams? 3 teams, each in a different org? What if I only observe the artifacts they produce vs. interview each of them in depth?

N != N (and N isn't the most important measure in case study research anyway)

- Most common critique when presenting case study research to engineers: "You only have 4 "N" how can you learn anything?
- Assumption: Researcher meant to use ***statistical sampling*** to achieve representative measure of population.
  - You might use statistical logic to choose your interviewees to inform on a particular case, but almost never to choose the cases you are comparing.
  - When you are **purposive sampling** (or selecting) achieving variation on your explanatory variables is what matters. General guidance: 4-10 is a good number.
- How do we select cases properly?

Case study selection logics

1. When it's ok to use a single case (see Yin 2009):
   - "Critical case" suitable to test predictions
   - Unique enough to warrant study regardless of generalizability.
   - Strong argument for representativeness
   - Longitudinal study enables comparison across time

- Otherwise:
  2. Analogy to experimental design (see Campbell and Stanley)
  3. Replication logic (See Eisenhardt 1989, Yin 2009)

- In all cases, you're choosing for theoretical reasons (e.g., how X explains/drives Y), reflected by RQ

2. "Quasi-experimental" design

- (Assuming familiarity with basic experimental designs)

Pretest-Posttest Control Group Design

~~R~~	O	X	O
~~R~~	O		O

Solomon Four-Group Design

~~R~~	O	X	O
~~R~~	O		O
~~R~~		X	O
~~R~~			O

Posttest-Only Control Group Design

~~R~~		X	O
~~R~~			O

R – Randomize

O – Observation (invasive)

X – Treatment (discrete)

2. "Quasi-experimental" design

Time Series

O O O O O O X O O O O

Equivalent Time Samples Design

✗ X_1O X_0O X_1O X_0O

Nonequivalent Control Group Design

<u>O X O</u>

O

Singe Case Study (extends to multiple)

O X O

Static-Group Comparison

<u>X O</u>

O

What you're looking for:

Know that "X" will happen (in the future). You start observing in advance, so you can watch how it changes things.

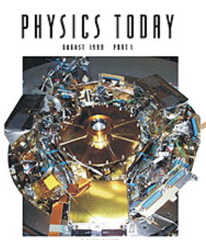
Advanced warning of X. Observe it happening, and find a similar group that it didn't happen to.

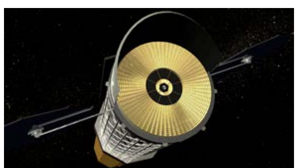
No advanced warning, but near identical group to compare to

Example: How does NASA tech funding model affect development process?

Big bang Model:

Spend 10+ years investing heavily in a **mission-enabling** capability that will likely only fly once.



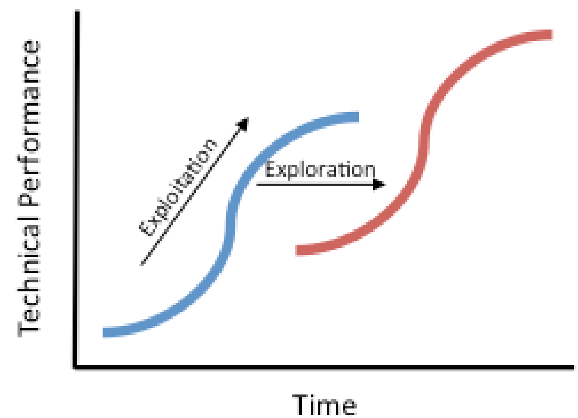


Example: X-ray Spectrometer for AXAF, initial development at NASA in 1983

Innovation theory says:

Inherently inefficient because the first build is always more expensive (on a per unit basis) and has lower performance than will future iterations. If the “2nd-nth” units are never produced, there will be

- no basis for averaging down R&D investment costs
- no benefit accrued from marginal production improvements.



Szajnfarber, Z. (2014) “Space science innovation: How mission sequencing interacts with technology policy” Space Policy 30(2) 83-90

Research Focus:

Merits of a few large missions vs. many small missions: Risk/reliability/survivability/tech obsolescence

http://www.darpa.mil/Our_Work/TTO/Programs/System_F6.aspx

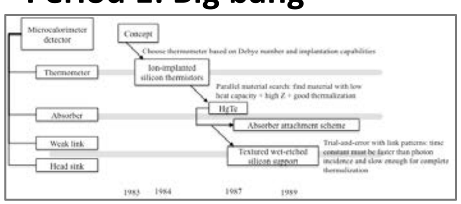
Quasi-experimental design:

Enabled by a unique empirical setting

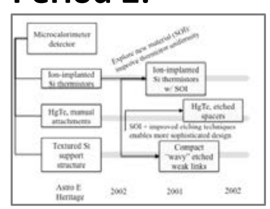
Mission context

R&D Periods

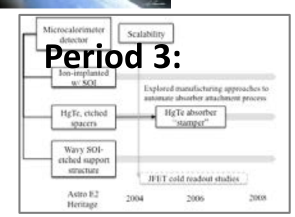
Period 1: Big bang



Period 2:



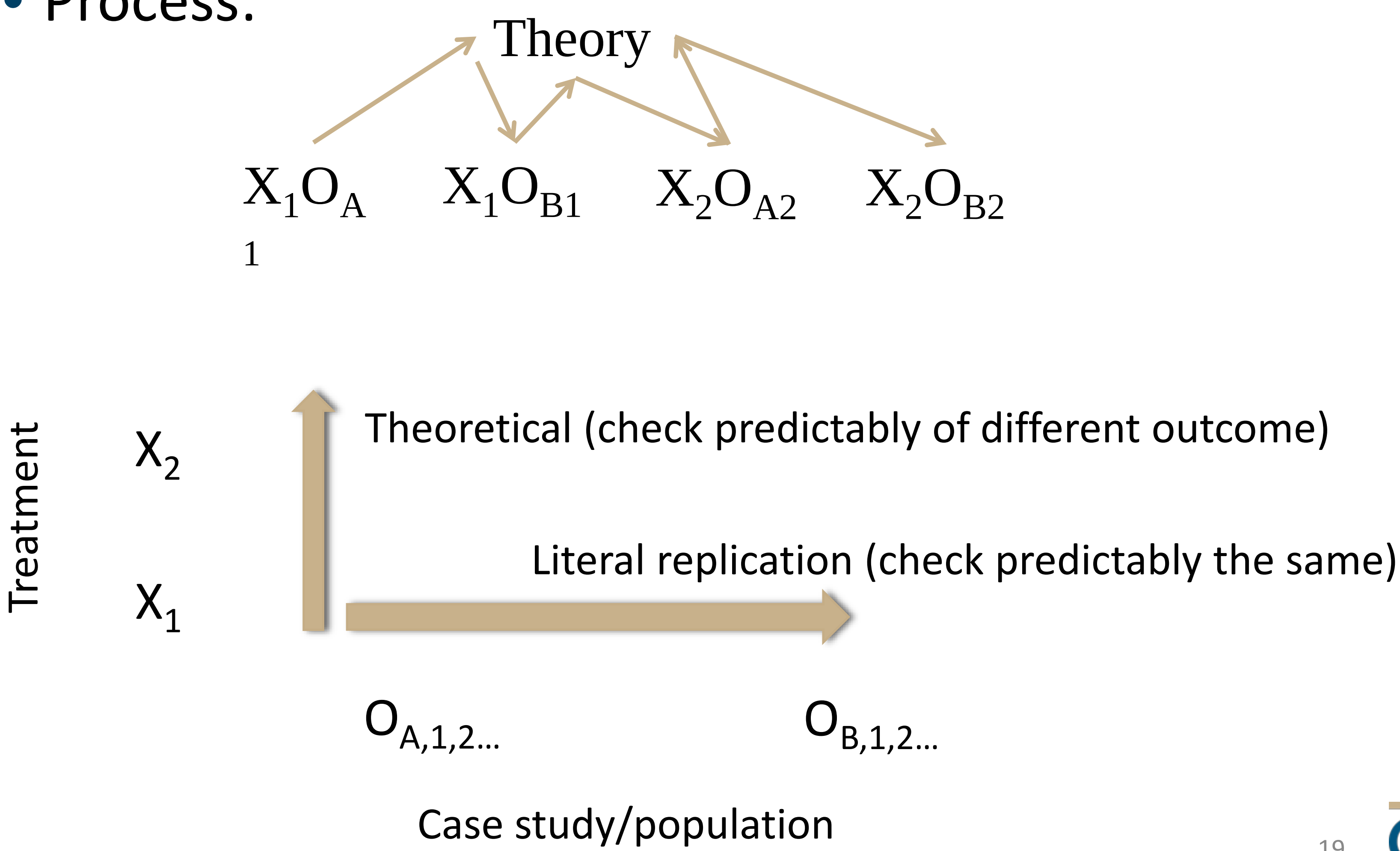
Period 3:



Rare insight into counterfactual: *what <u>would</u> have happened <u>if</u> the mission opportunities had been structured differently?*

3. Replication Logic

- Process:



Summary

- Analogy to experiments: quasi-experiments
  - Choose cases to be able to rule out alternative explanations of the observed effect.
- Replication logic:
  - Progressively gain confidence in ability for emerging theory to make predictions under different conditions.

Relating back to validity (I and E)

- Internal Validity (necessary minimum):
  - Level 0: Are you in fact observing the phenomenon you think you are?
    - Easiest to guarantee in qualitative case studies. Hard with other methods.
  - Level 1: Can you isolate the impact (causality) of the **treatment** in your observations?
    - Largely done through selection of cases/depth of observation
- External Validity (asks the question of generalizability):
  - Level 0: Is the **effect** repeatable in all contexts of this kind?
    - This is the value of doing at least one literal replication
  - Level 1: How broadly does it apply: To what populations, settings, treatment variables, and measurement variables can this effect be

Qualitative case studies may (are capable of) generalize farther than quantitative ones, with good selection of cases and supporting data.

GW

Side note:

On selecting informants/who to observe

- Here, you are aiming to be representative of the case
- Sampling:
  - Non-probability sampling:
    a. **Purposive (judgmental) sampling:** The units to be observed are selected on the basis of the researcher's judgment about which ones will be the most useful or representative. (Appropriate for small N)
    b. **Snowball sampling:** each person interviewed may be asked to suggest additional people for interviewing.
    c. **Quota Sampling:** Units are selected into a sample on the basis of prespecified characteristics, so that the total sample will have the same distribution of characteristics assumed to exist in the population being studied.
  - **Probability Sampling**: The general term for samples selected in accord with probability theory, typically involving some random-selection mechanism.
    a. **Equal Probability of Selection Method**: A sample design in which each member of a population has the same chance of being selected into the sample.
    b. **Simple Random Sampling**: A type of probability sampling in which the units composing a population are assigned numbers. A set of numbers are then

Scoping and conducting data collection

Data Type	Description	Strength	Weakness	Appropriate Use
Documentation	Written documents produced in normal operations (e.g., e-mail, calendars and meeting minutes, proposals, status updates, reports)	Readily available, often stored in searchable formats Near real-time source of information	Can be incomplete and quite biased. Nearly impossible to determine direction of bias	Most useful to structure/focus interview questions on particular issues and then later to corroborate evidence from other sources
Archival Records	A d… officially published (e.g., budget or personnel records)	… Tends to be complete if it exists and aggregates large quantities of data	… than informal documents.	… plots in the final write-up. Generally not used to build theory.
Interviews	Refers to in person questions and answers with an info…	The only way to directly probe the "whys" of the …	Quality of information gained can be highly variable, due to …	Key part of most qualitative studies, but make sure to … ne … biased sources.
Direct Observation	Real-time observations of the phenomenon as it unfolds	Unique lens into the process, in context. Enables real … unfiltered view of actions.	Inherent limits in scope of what can be observed can drive … phenomena	Use when possible. Can reduce scope of observation by focusing … simulation exercises).
Physical Artifacts	A physical object produced during/by the phenomenon (e.g., posters and mission patches, the system)	Can represent externalization of cultural values (less common in systems engineering and design studies). May enable evaluation of "performance;" for example, the performance of a system produced by a design process.		Can complement other sources

Important for cross-checking e.g., interview responses.

Let's you get in designer's head, must be done retrospectively. Some phenomena take to long to observe (or can't be)

Let's you see phenomenon evolve in real-time, limits to what you can reasonably observe.

See paper for tips and tricks

Inductive Analysis Strategies

Avoiding “death by data asphyxiation”

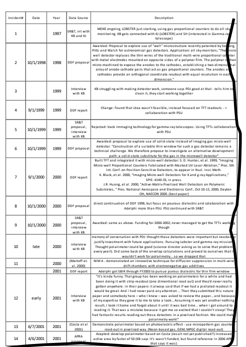
Process Data	Within-case “sense-making”	Cross-case theory building

~100 hrs interviews

~150 archival documents

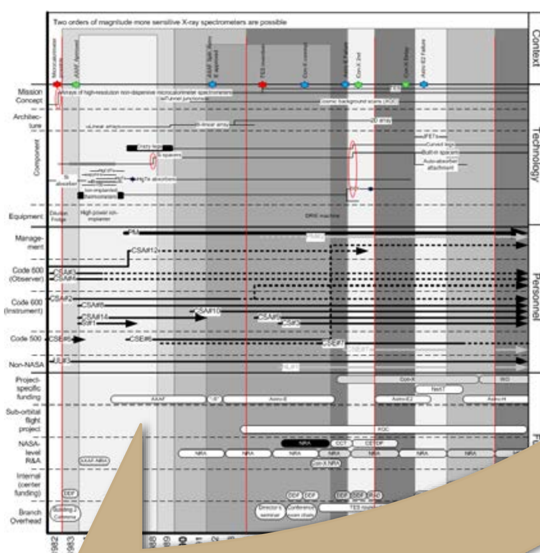
~2 months informal observation

Analytical Chronologies (Pettigrew 1990)

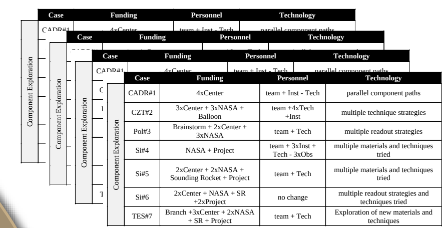


Event Database (Van de Ven et al 1990; 2000)

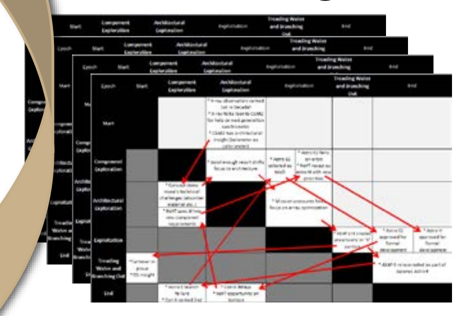
Structured Visual Ma (per Langley 1999)



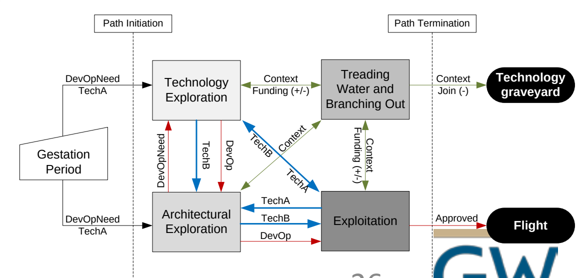
Characteristic Epochs



T nsition inducing Shocks



Epoch-Shock Model



Avoiding "death by data asphyxiation"

Process Data	Within-case "sense-making"	Cross-case theory building

Analytical Chronologies (Pettigrew 1990)

~100 hrs inter…

~1… arc… docu…

~2 months informal observation

Event Database (Van de Ven et al 1990; 2000)

Structured Visual Ma… (per Langley 1999)

Characteristic Epochs

Case study myth:

Not enough data
Actually, often more data than you know what to do with.

Engineering discomfort:

Despite many textbooks on process of inductive research, no method that spits out a weight and a p-value.

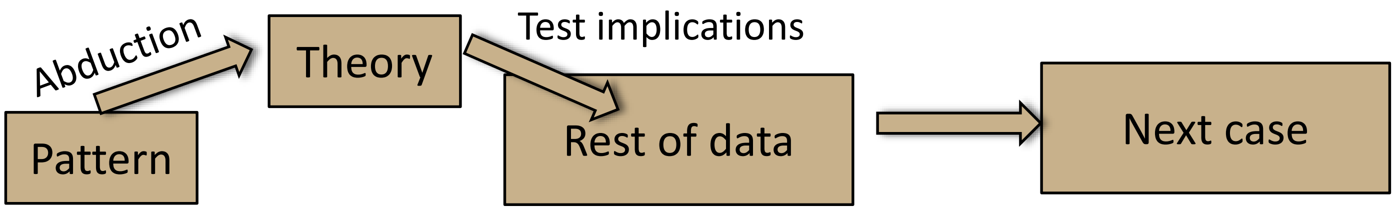
Need for "creative leap"

Abduction (the creative leap)

- Abductive reasoning:
  - Inferring ***a*** as an explanation of ***b***. ***B*** is the consequence (or observed outcome) and ***a*** is the abducted (ideally best) explanation.
  - ***A*** is not guaranteed to be true (simply by this abduction), but the validity can then be tested deductively.
- Abductive steps show up in most research even though they are often not acknowledged (e.g., where do hypotheses come from?)

Avoiding (the bad kind of) Bias

- How do we make sure that an insight from a small number of (e.g., interview-based) case studies is true?
  - Often asked: Would multiple people looking at the same data come to the same conclusion?
    - Analogy to repeatability (incorrect logic)/inter coder reliability
  - Better question: How can I prove that my abduced explanation fits the data?
    - Analogy to training data

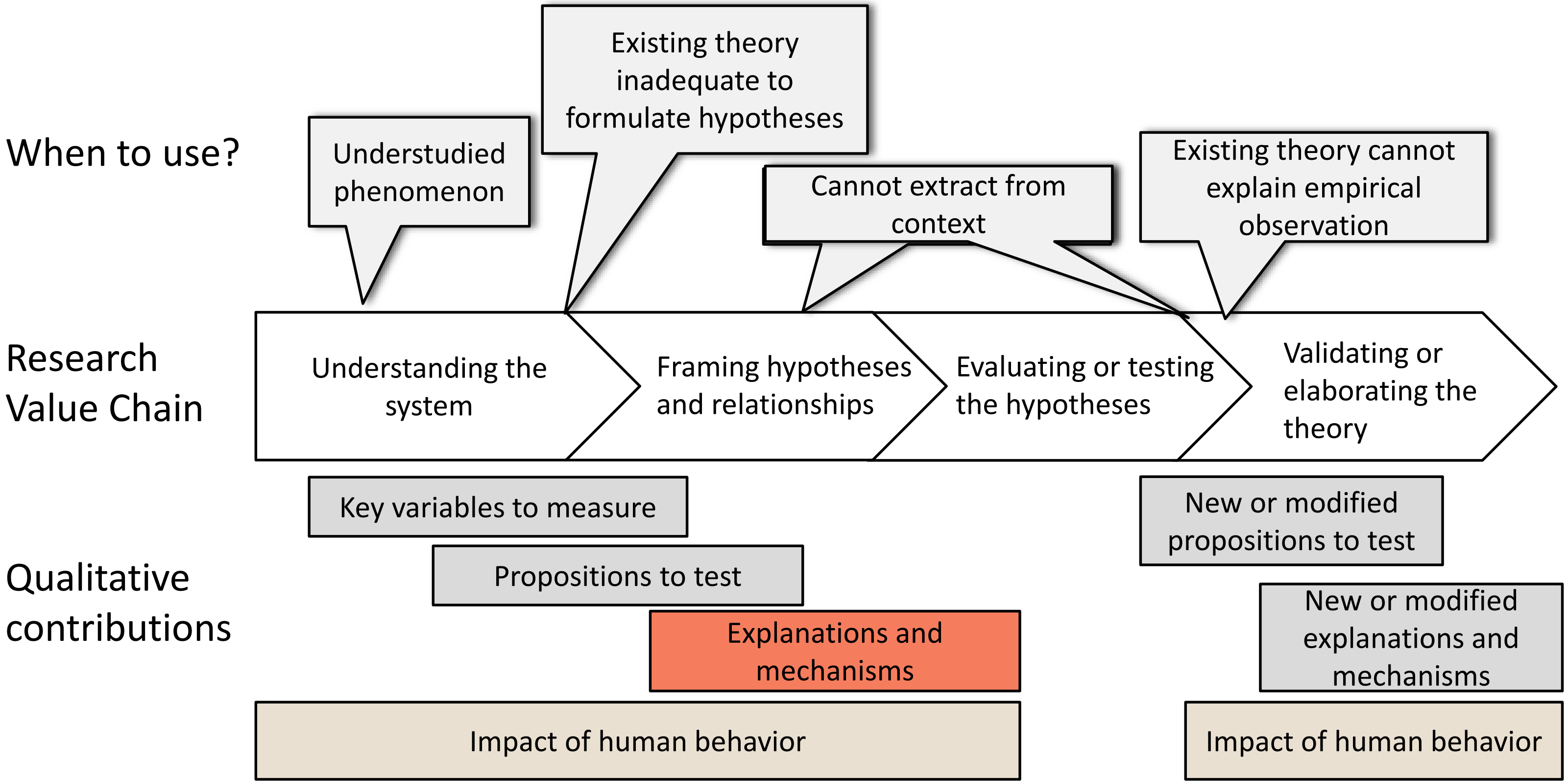


- Key point: It doesn't matter if multiple people could come up with the insight. It is critical that the validity of the insight can be objectively proven.

Example: Why do technology development paths appear to "switchback"?

- Began to see to explanatory "patterns" coming up over and over again in my first two instances.
- I checked whether they explained what the observation in several other instances (selected using replication logic) and they did, but there was also a third different reason.
- Tried the 3 on two more cases (again, replication logic) and they explained the observations and no new "patterns" emerged.
- Stopped at "theoretical saturation"

Output of case studies



Case studies rarely "prove" anything. They help us deeply understand how a process or phenomenon works. This is the building block for future theory or a way to elaborate existing theory.

How should you judge if a case study result is good/valid?

Qualitative approaches: when and why?

- Why use qualitative research approaches?
  - Study socio-technical systems: messy complexity of human and organizational drivers of design, development, operation
- When to use qualitative research approaches?
  - When the phenomenon is not easily observable or quantifiable, e.g. occurs inside the minds of actors
  - When existing theory is inadequate to explain the phenomenon
    - Perhaps because theory derived in a different context, or disproved by empirical evidence, or not investigated empirically.
    - Might be manifested as inability to come up with hypotheses, not clear what to measure, not enough knowledge to make good modeling assumptions
  - When the phenomenon must be studied in empirical context
    - Perhaps because impractical to replicate in laboratory or model, empirical details too important to abstract away [e.g. disaster response decision-making]

Standards for evaluating case studies

- Caution: different process, different standards

1. Were the cases picked to enable inference that answers the posed questions?
   - Check selection, replication logic
   - Don’t sample on the dependent variable (don’t choose because the outcomes are different)
   - Strong theoretical grounding is critical
2. Do the data fit the proposed explanation?
   - Were alternative explanations explored and ruled out?
   - Did they talk about saturation on theoretical dimensions?
   - Did they take advantage of depth?
3. Is the evidence compelling as written?
   - Balance “showing” the data and “telling” the findings
   - Do not seek objectivity at the expense of unique insight
   - “Plausibly Generalizeable” is enough.

Further Reading

- Our paper: “Qualitative Methods for Engineering Systems: Why we need them and how to use them” (in review) (joint with E. Gralla) – provided read-ahead
- Edmondson, A. C., & McManus, S. E. (2007). Methodological fit in management field research. *Academy of Management Review*, *32*(4), 1155–1179. https://doi.org/10.5465/AMR.2007.26586086
- Eisenhardt, K. M. (1989a). Building Theories from Case Research. *The Academy of Management Review*, *14*(4), 532–550.
- Eisenhardt, K. M. (1989b). Building Theories from Case Study Research. *Academy of Management Review*, *14*(4), 532–550.
- Langley, A. (1999). Strategies for Theorizing from Process Data. *Academy of Management Review*, *24*(4), 691–710. https://doi.org/10.5465/AMR.1999.2553248
- Locke, K. (2001). *Grounded theory in management research*. Sage.
- Mintzberg, H. (1979b). An emerging strategy of“ direct” research. *Administrative Science Quarterly*, 582–589.
- Pettigrew, A. M. (1990). Longitudinal Field Research on Change: Theory and Practice. *Organization Science*, *1*(3), 267–292.