

<https://github.com/margaretstorey/cbsoft2019tutorial/>

A Tutorial: **Studying Human and Social Aspects in Software Engineering** *(My Publish or Perish Keynote: The Uncut Version)*

SBES 2019, Salvador, Brazil



Margaret-Anne Storey
@margaretstorey



University
of Victoria





“all models are wrong, but some are useful”

– Box, 1976

Motivation

Paradigms (activity)

Methods (activity)

Contributions (activity)





Motivation

Paradigms (activity)

Methods (activity)

Contributions (activity)

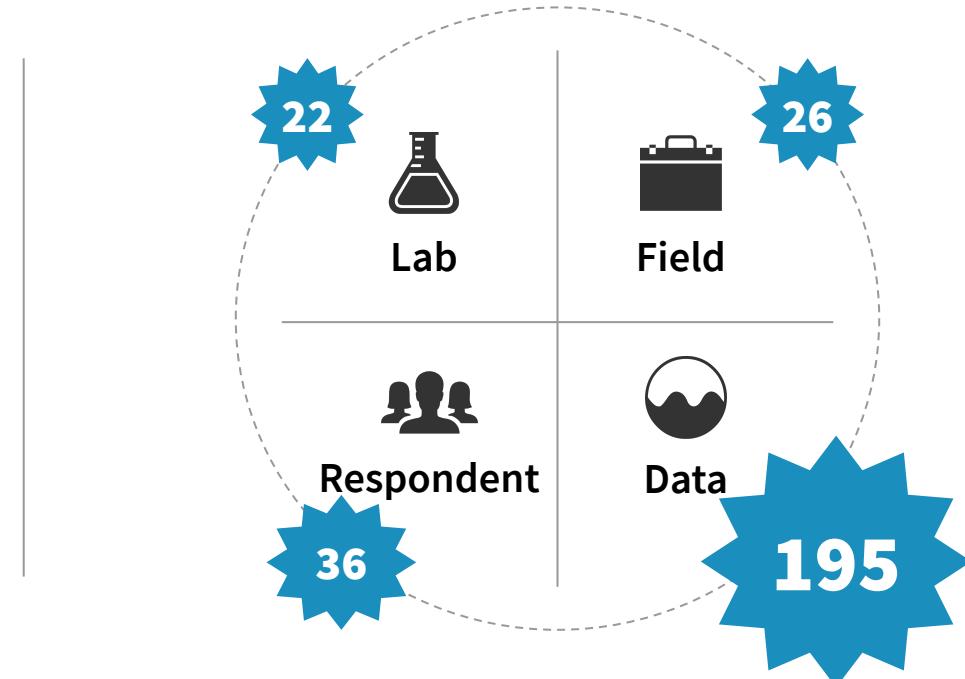
Adds to our knowledge of tools, processes and practices

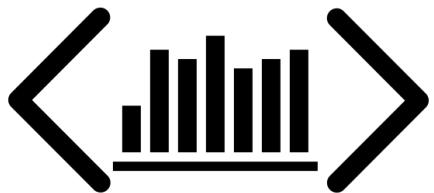
Addresses **gaps** in knowledge

Replicates knowledge by testing old knowledge with new participants or new research sites

Expands knowledge by extending research to new ideas or practices

Broadens our perspectives – e.g. add voices of individuals to the body of knowledge
Inform practice by developing new ideas





“Measurement is the empirical, objective assignment of numbers, according to a rule derived from a model or theory, to attributes of objects or events with the intent of describing them.”

– Kaner, 2004

Product Metrics:

KLOC, Complexity measures (cyclomatic complexity, function points), OO metrics, #defects

Field metrics:

User engagement, user sentiment

Process metrics:

Testing, code review, deployment, agile practices (e.g., #sprints, burndown rate)

Productivity:

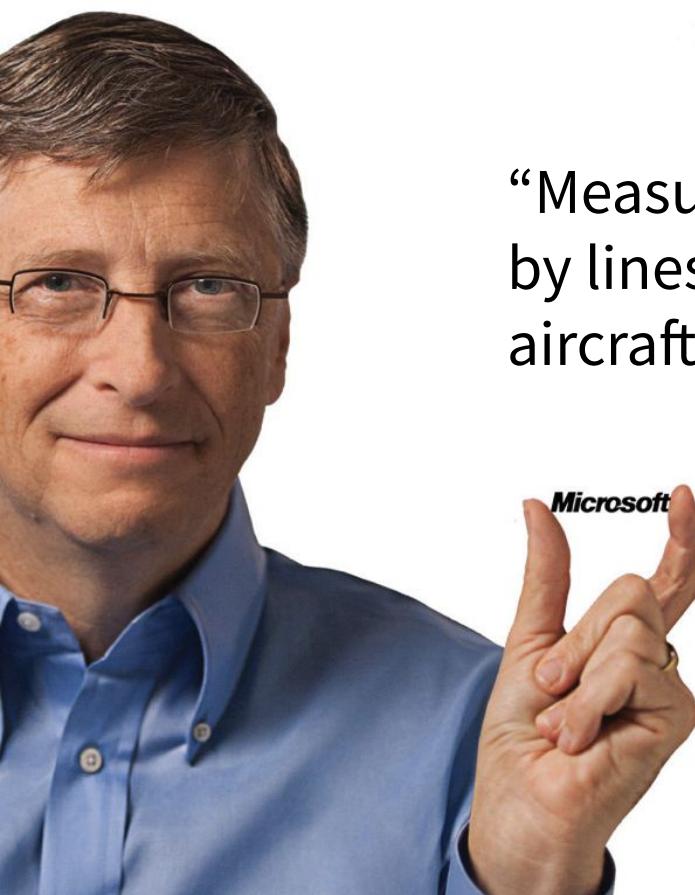
KLOC, Mean time to repair, #commits, team sprint velocity

Developer metrics:

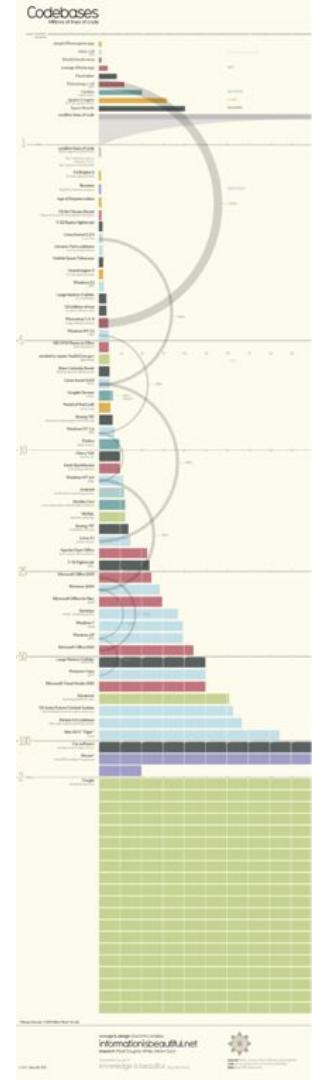
Skills, followers, biometrics

Estimation:

cost metrics and models



“Measuring programming progress by lines of code is like measuring aircraft building progress by weight.”



Program data:

runtime traces,
program logs, system
events, failure logs,
performance logs,
continuous
deployment,...

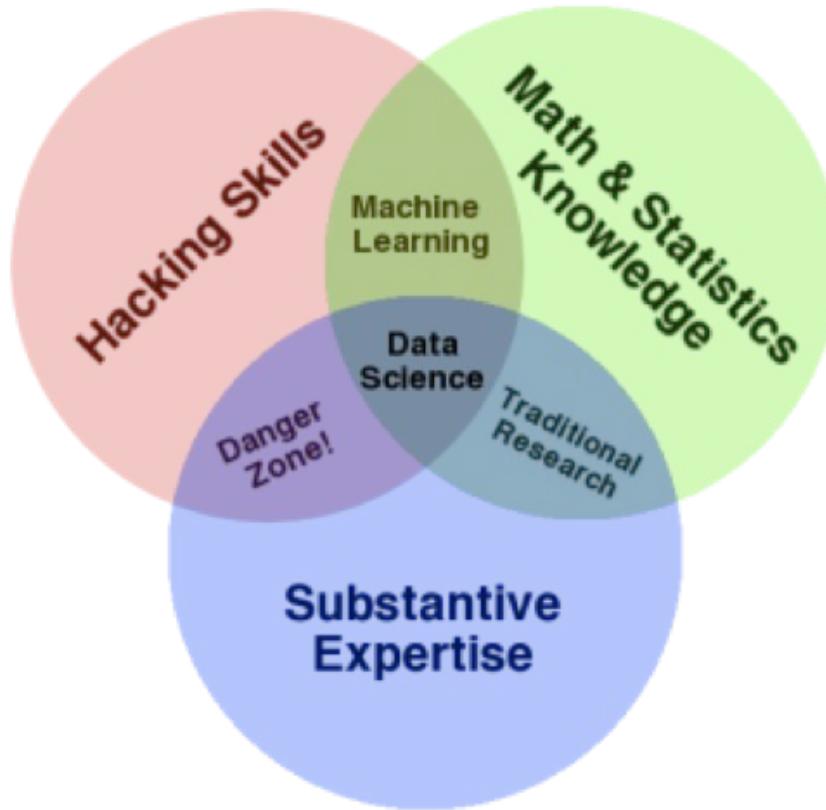
User data:

usage logs, user
surveys, user forums,
A/B testing, Twitter,
blogs, ...

Development data:

source code versions,
bug data, check-in
information, test
cases and results,
communication
between developers,
social media







Ownership
Churn
Tangled code changes

Poor replication
Poor actionability
“Secret life of bugs”

Data

Data may have low
construct validity

Data assumes humans
are “rational animals”

Data does not tell you
why

Perils from using GitHub data:

A repository is not necessarily a (development) project

Most projects are inactive or have few commits

Most projects are for personal use only

Only 10% of projects use pull requests

History can be rewritten on GitHub

A lot happens outside of GitHub

Data

Data may have low construct validity

Data assumes humans are “rational animals”

Data does not tell you why

Analysis

Correlations are not cause and effect

Big data and small effects

Researcher bias

Martin Shepherd's meta analysis of 24 studies defect prediction:

Technique used had a small effect

but

Research group that did the study had a bigger effect

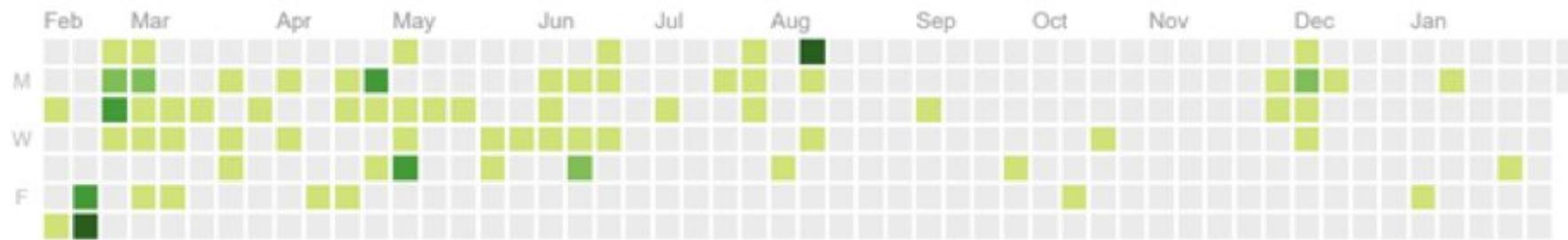
Data	Analysis	Aftermath
Data may have low construct validity	Correlations are not cause and effect	Low actionability Ethics of using data
Data assumes humans are “rational animals”	Big data and small effects	Unexpected consequences
Data does not tell you why	Researcher bias	Biases in algorithms (feedback loops)

“Defect prediction approaches are evaluated on the past history of a system’s bugs, where that history is treated as the future. A real prediction perturbs the space–time continuum. Without real world adoption, you simply can’t measure the predictor’s effect. A real prediction perturbs the space–time continuum.”

– [Lanza et al., 2016](#)



Contributions



Summary of pull requests, issues opened, and commits. [Learn how we count contributions.](#)

Less More

Contributions in the last year

235 total

Feb 8, 2015 – Feb 8, 2016

Longest streak

6 days

February 20 – February 25

Current streak

0 days

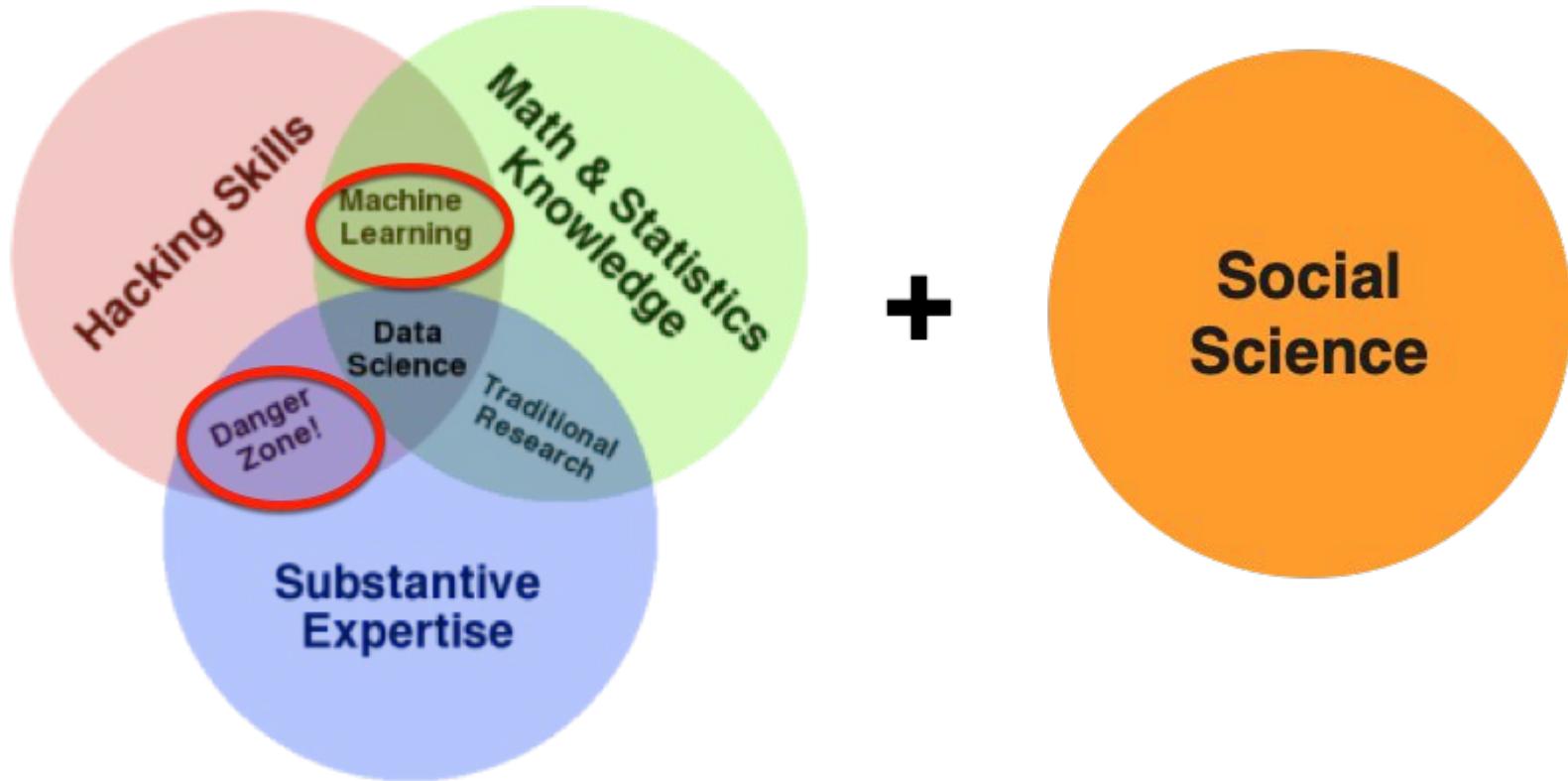
Last contributed 12 days ago

19 | “Contributing graphs considered harmful”, Scott Hanselman



20 | Biases in algorithms and diversity

<https://www.sciencemag.org/news/2017/04/even-artificial-intelligence-can-acquire-biases-against-race-and-gender>

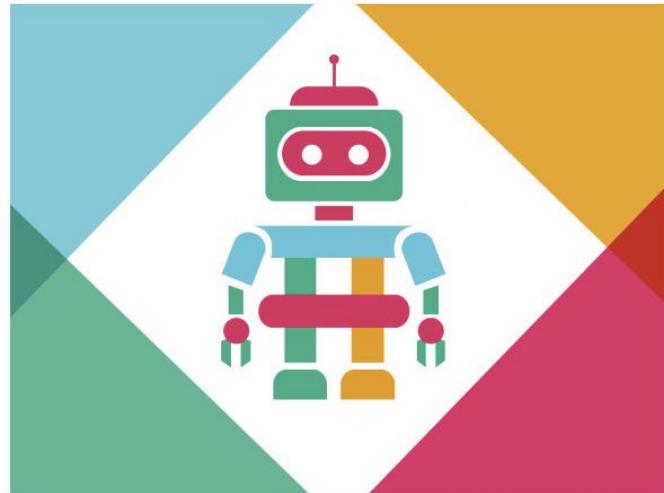


21 |

“Most big data is social data – the analytics need serious interrogation”

<http://blogs.lse.ac.uk/impactofsocialsciences/2015/02/12/philosophy-of-data-science-emma-uprichard/>

SLACK IS OVERRUN WITH BOTS. FRIENDLY, WONDERFUL BOTS



“You’ll be paid in the future based on how well you work with robots.”

– **Kevin Kelly, Futurist**

Lies, damned lies and software analytics, why big data needs thick data, Storey, 2016

Mad about Measurement, De Marco

The Emerging Role of Data Scientists on Software Development Team, Miryung Kim, Thomas Zimmermann, Robert DeLine, and Andrew Begel, ICSE 2016.

Analyze This! 145 Questions for Data Scientists in Software Engineering, Andrew Begel & Thomas Zimmermann, ICSE 2014

Software Analytics Tutorial, Dongmei Zhang & Tao Xie, ICSE 2012

Rules of Data Science in SE, Tim Menzies

Researcher bias: The use of machine learning in software defect prediction, Sheppard et al., IEEE TSE 2014.

Eight (No, Nine!) Problems with Big Data", Gary Marcus and Ernest Davis, New York Times, April 6, 2014.

Dwelling in Software: Aspects of the felt-life of engineers in large software projects, Harper et al., ECSCW 2013. Springer.

The secret life of bugs: Going past the errors and omissions in software repositories. Jorge Aranda & Gina Venolia, ICSE 2009.

Algorithms in our Midst: Information, Power and Choice when Software is Everywhere, Zeynep Tufekci, CSCW 2015. , ACM..

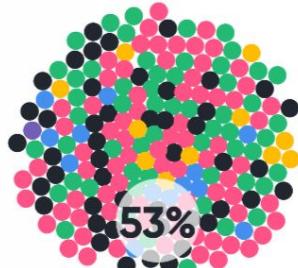
Test coverage and post-verification defects: A multiple case study, Audris Mockus, Nachiappan Nagappan, Trung T. Dinh-Trong, ESEM 2009

<https://github.com/margaretstorey/cbsoft2019tutorial/>



Paradigms Activity: Name badge

What is your predominant worldview?



Postpositivist



Constructivist



Advocacy
researcher

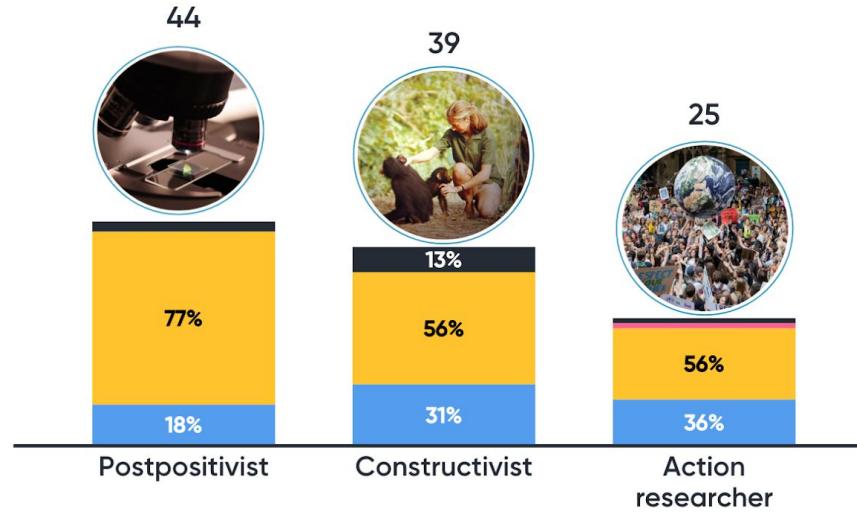


Role?

- Practitioner
- Industry researcher
- Student
- Researcher (faculty, postdoc)
- Other
- Unknown

392

What is your predominant worldview?



Role?

- Student
- Researcher (faculty, postdoc)
- Industry researcher
- Industry practitioner
- Other
- Unknown

Screenshot

108

Name

Research area

Research paradigm
(if you know)

Postpositivist



Constructivist



Pragmatist



Advocate





Peggy "Constructivist"

Motivation

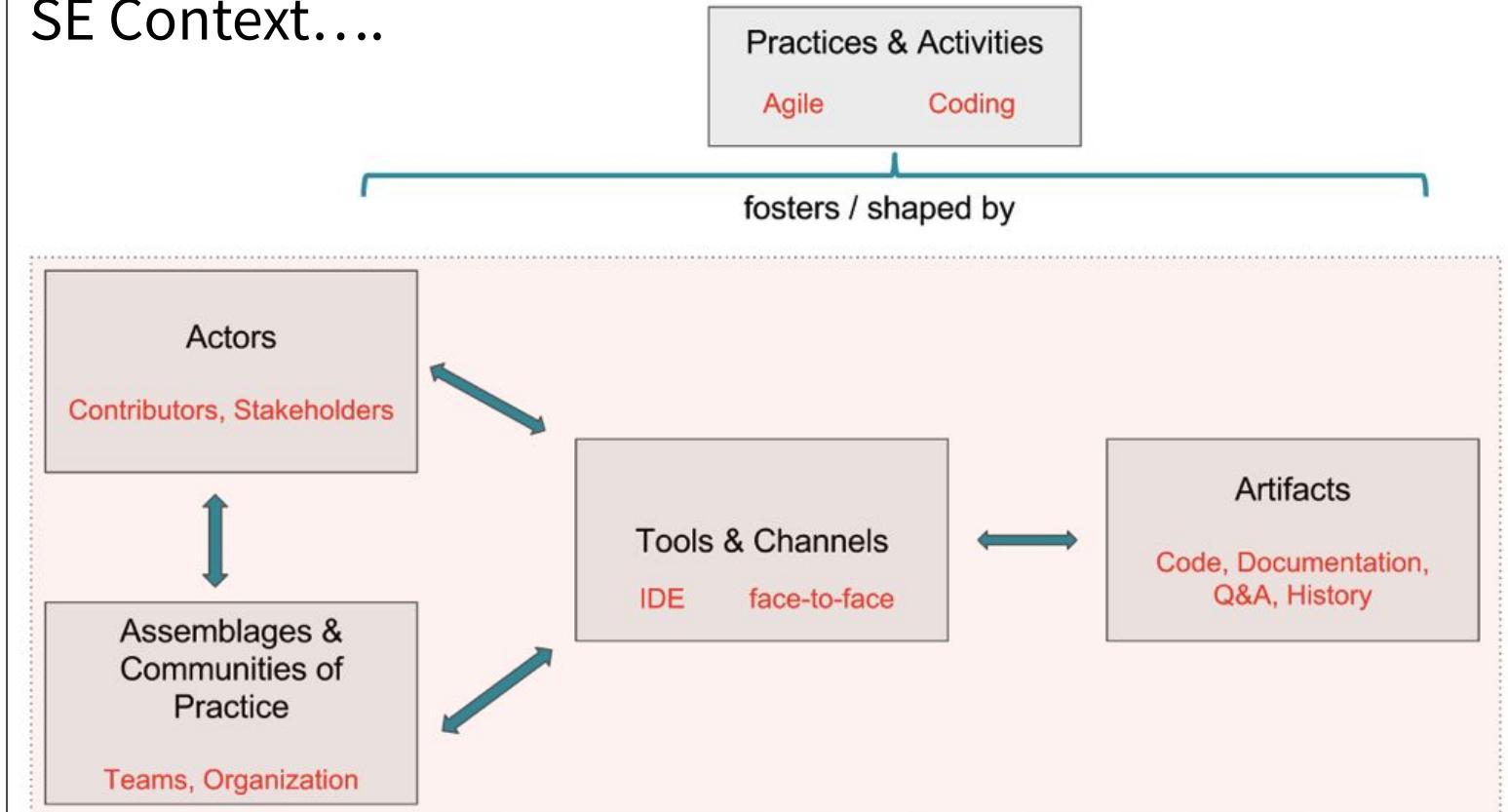
Paradigms (activity)

 **Methods (activity)**

Contributions (activity)

Actors:	Behavior:	Context:
human systems, individuals, groups, organizations, communities	all aspects of the states and actions of those human systems	temporal, locational and situational features in which the human system is embedded

SE Context....

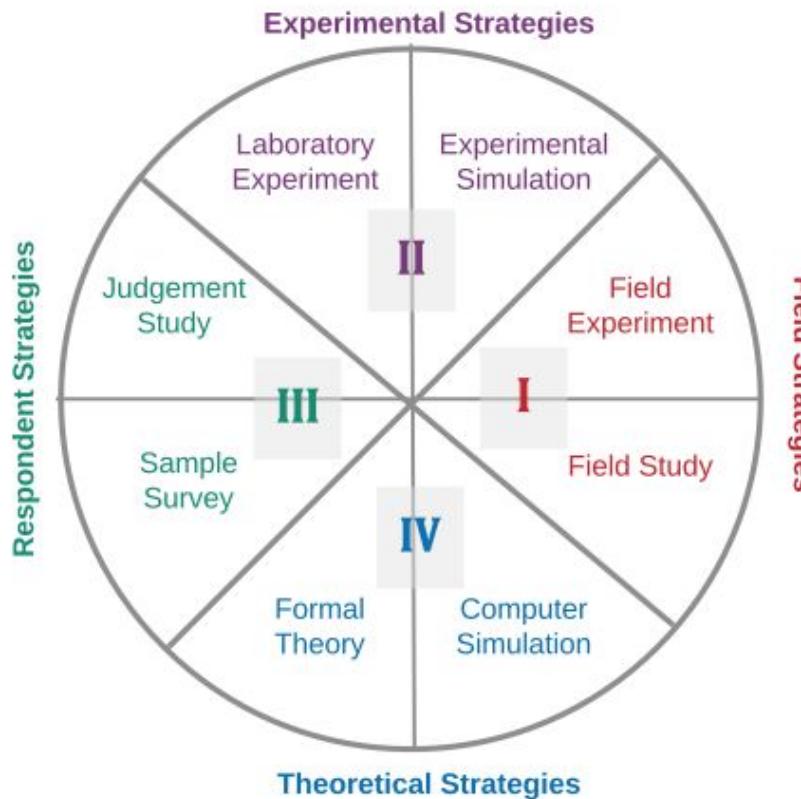


31 | Who, what, where?

Some **content** that is of interest

Some **ideas** that give meaning to that content

Some **techniques** or procedures for studying the content and ideas



33 | McGrath's Circumflex

Joseph. E. McGrath.
Methodology matters: Doing research in the behavioral and social sciences. 1972

Generalizability

of the evidence over
the populations of
actors

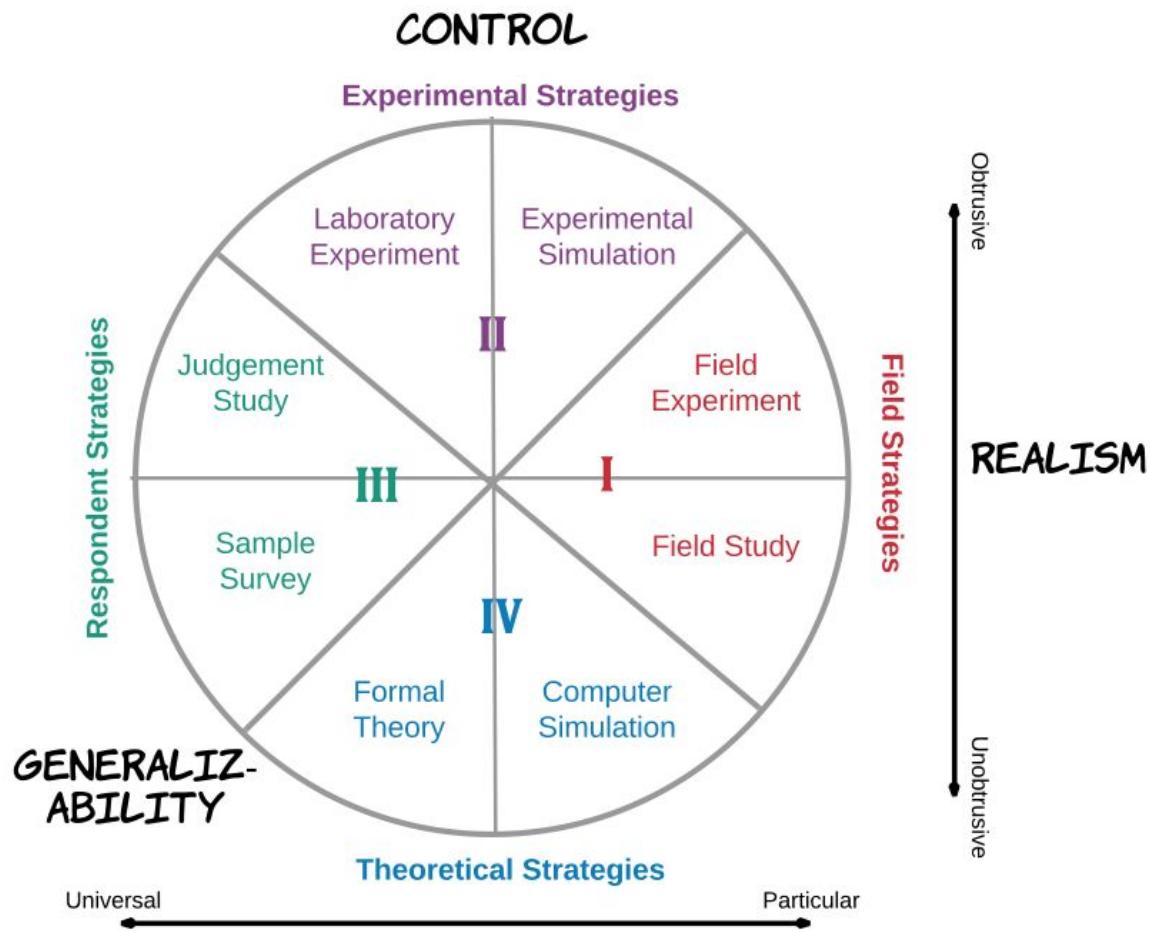
Precision (control)

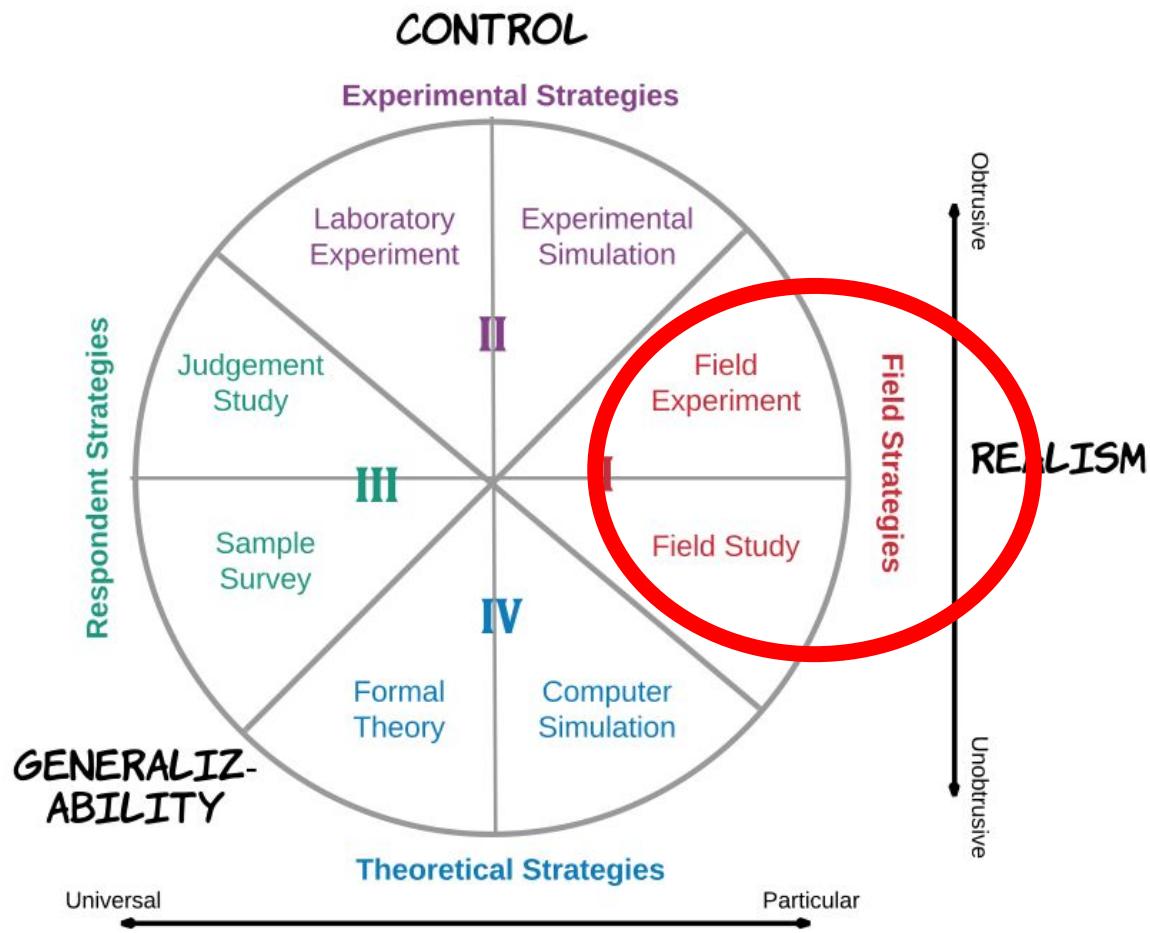
of measurement of
the human
behaviours being
studied

Realism

of the situation or
context where the
evidence is gathered







Field studies:

Ethnography

Case studies of organizations

Field experiments:

Naturalness is given up for increased measurement precision

Some variable may be manipulated (e.g. tool used or process)

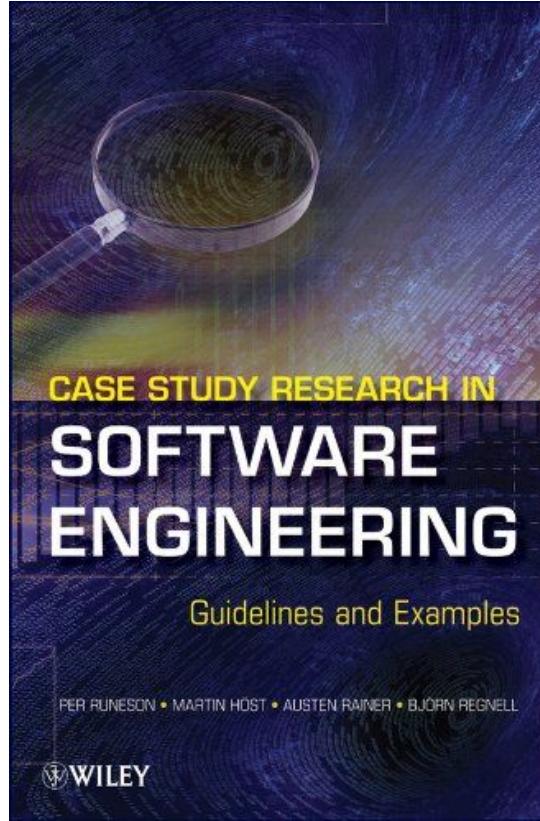
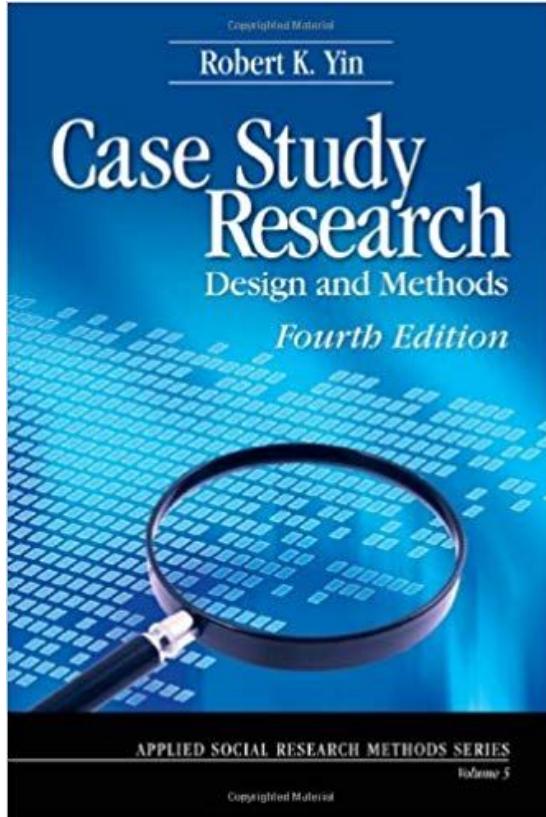
Highly obtrusive



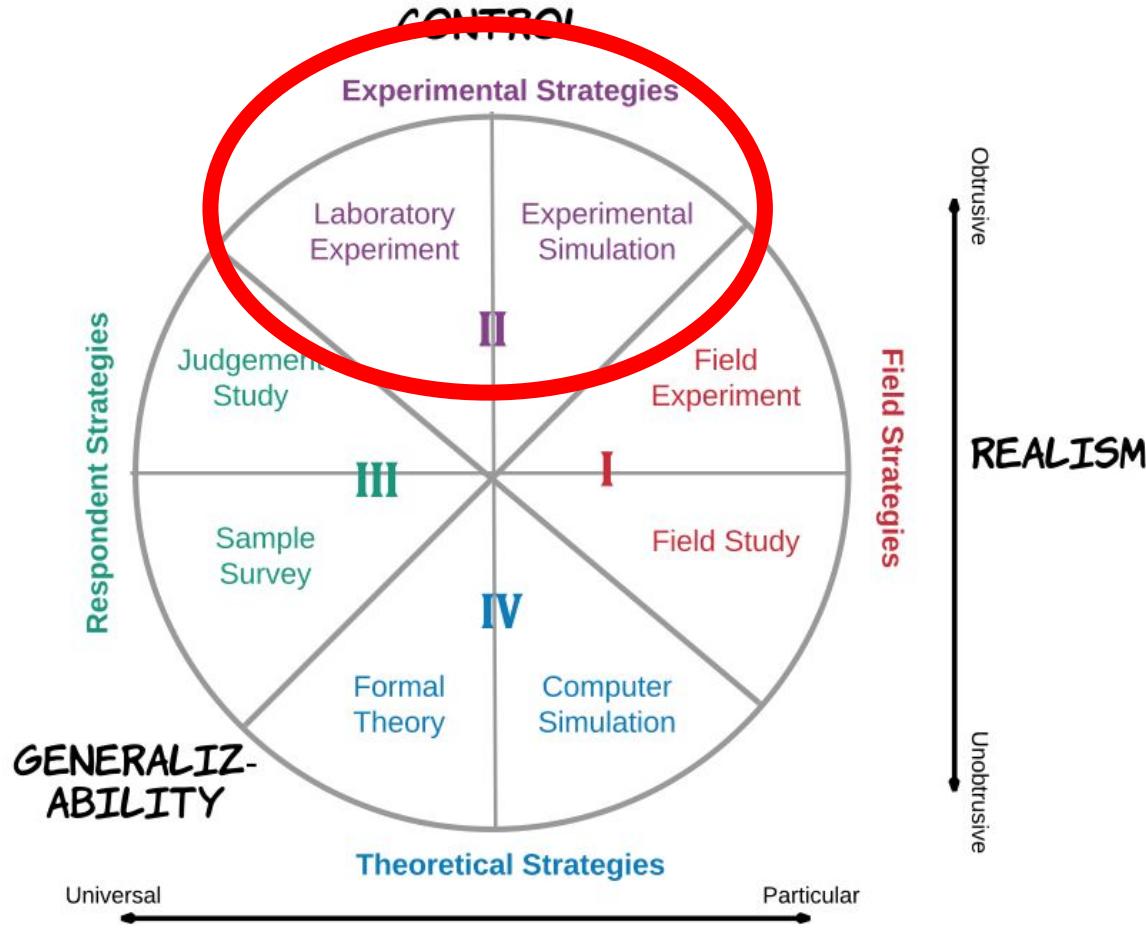
"If you want to understand what motivates a guy to pick up skateboarding, you could bring him into a sterile laboratory and interrogate him... or you could spend a week in a skatepark observing him interacting with his friends, practicing new skills and having fun."

Ethnography is observing people's behavior in their own environments so you can get a holistic understanding of their world—one that you can intuit on a deeply personal level."

—LiAnne Yu, cultural anthropologist



39 | Case studies in software engineering



Laboratory experiment:

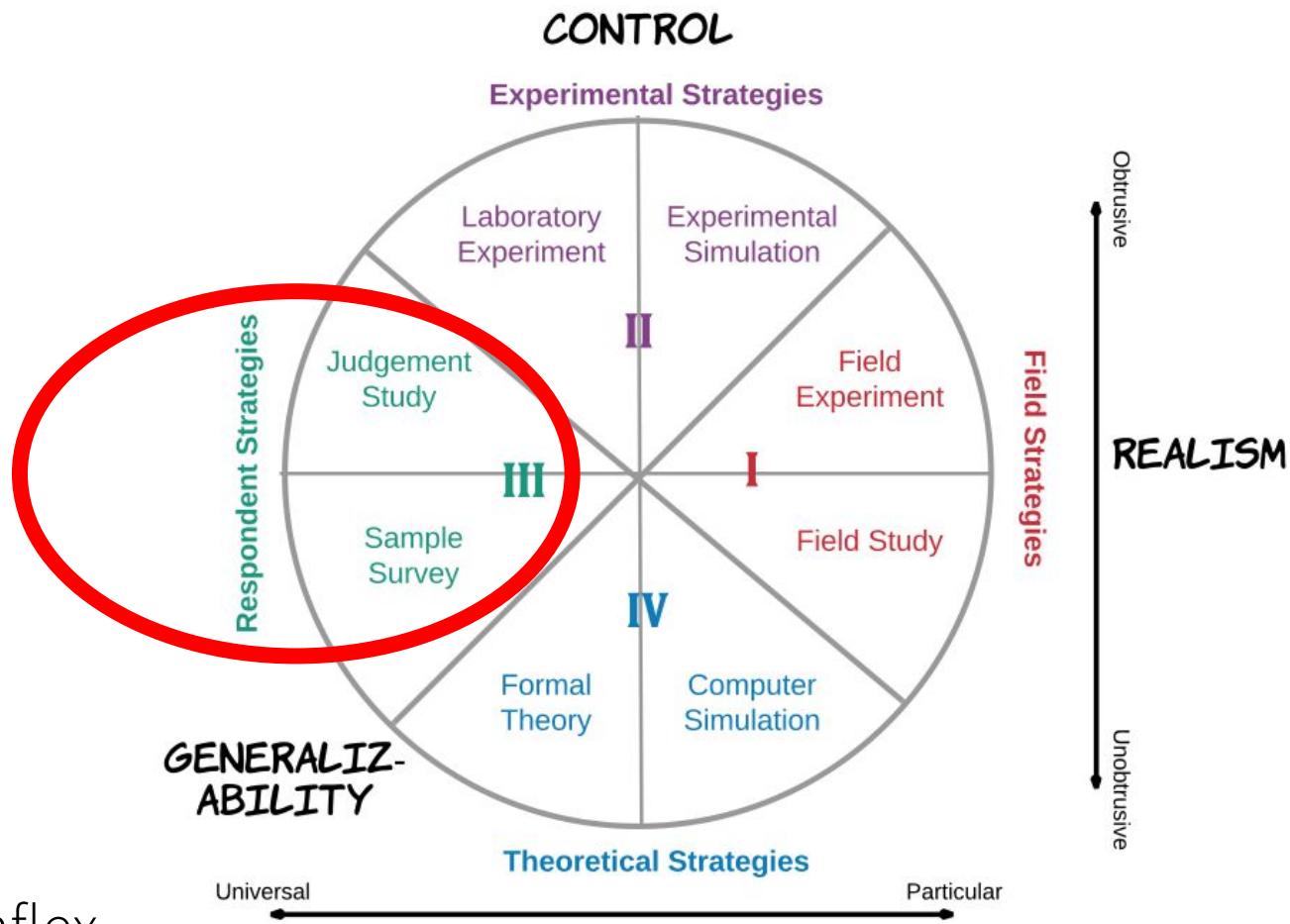
Investigator created setting,
defined rules for its operation,
induces human actors to participate

Increased precision of measurement
over human behaviour

Increased obtrusiveness, unrealistic
setting and reduced generalizability

Experimental simulation:

Experimenter has control over the
setting and conditions – but feels
more like the real setting
e.g. flight simulators



Sample survey:

Collects evidence across a distribution of some variables or relationships among them, within a specific human actor population

Careful sampling must be done to maximize generalizability

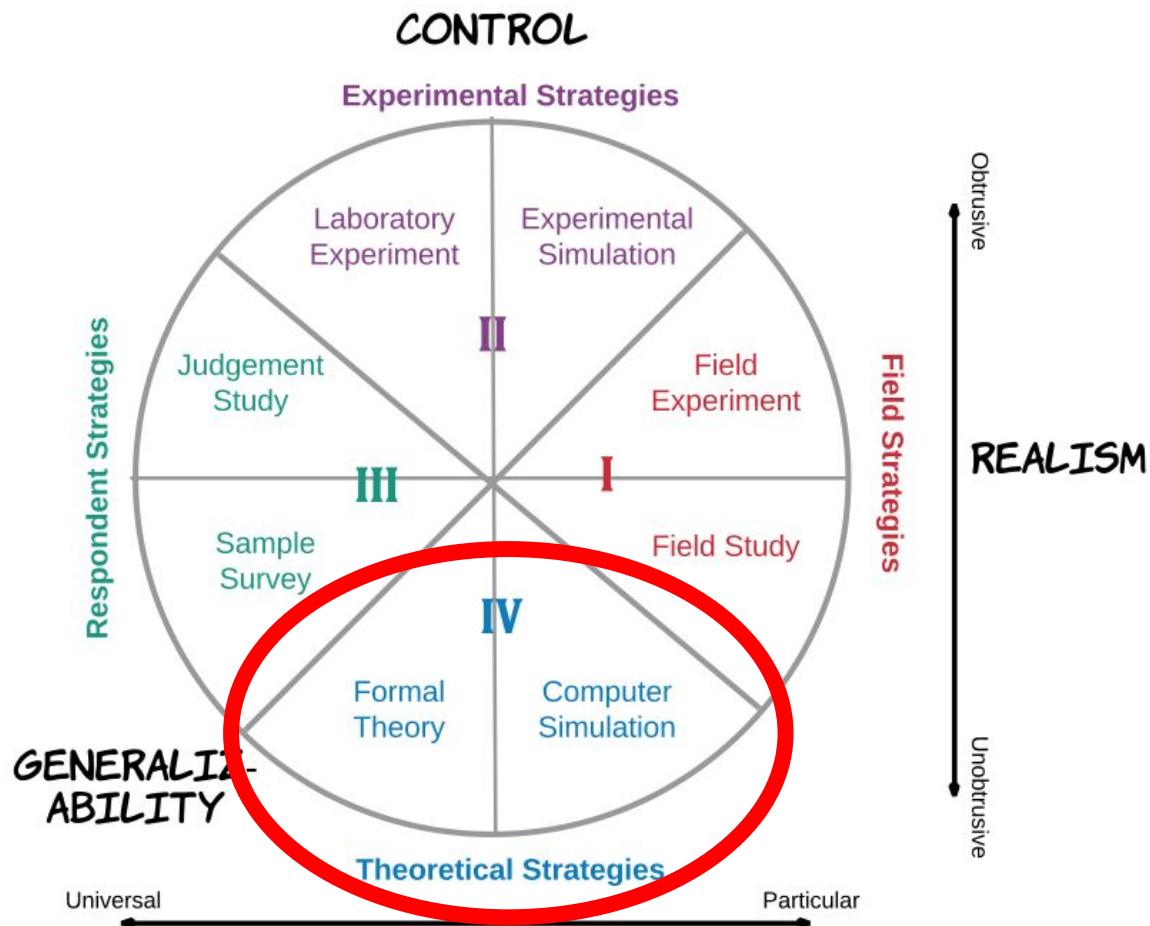
Little opportunity for much precision of measurements

Judgment study:

Obtaining information about a set of stimulus materials

Usually done with actors of convenience

More precise measurements, but low generalizability



Formal theory:

Theories based on previous empirical evidence or theories

No gathering of empirical observations – relations among variables of interest are formulated

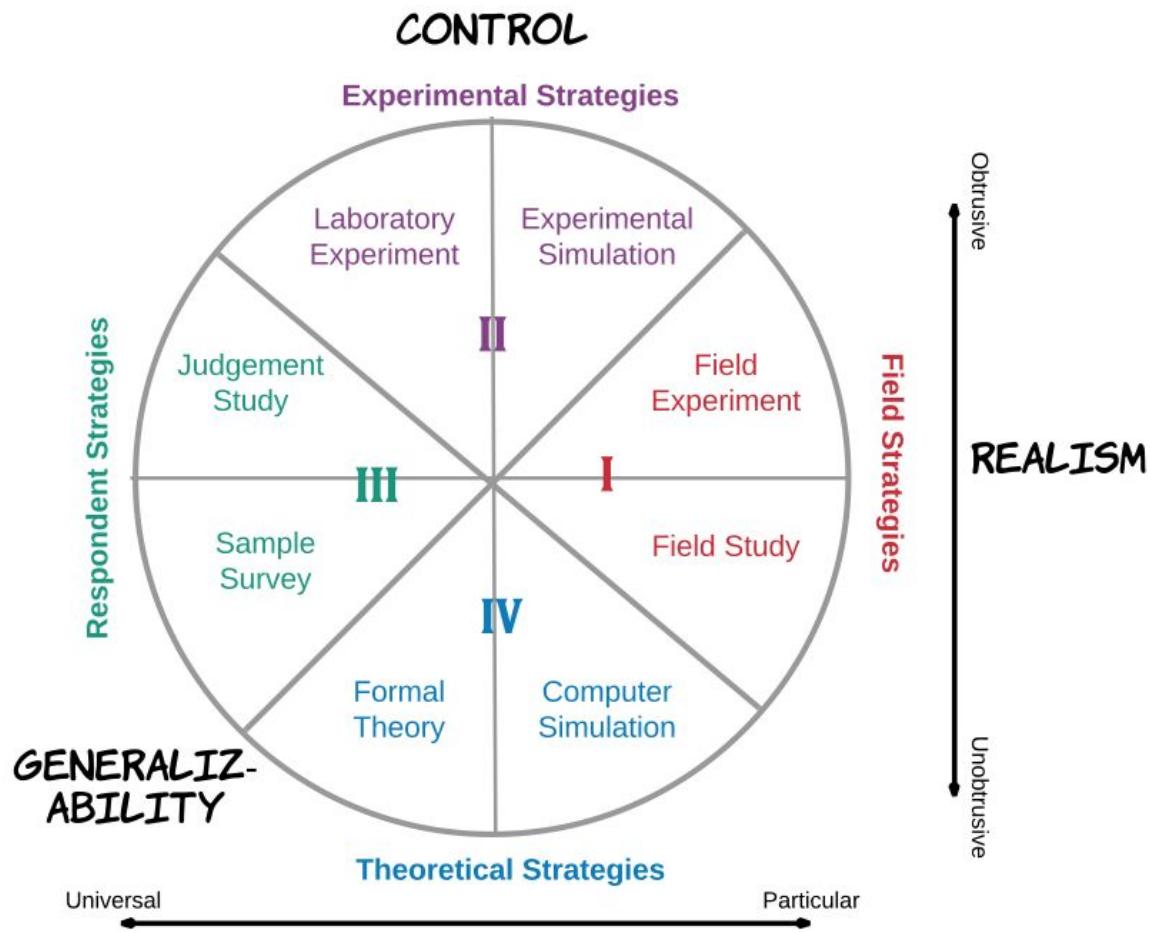
Relations (hypotheses, propositions) should hold over a broad range of populations

Computer simulation:

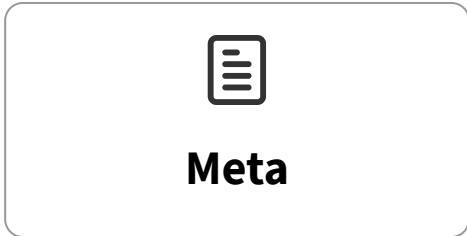
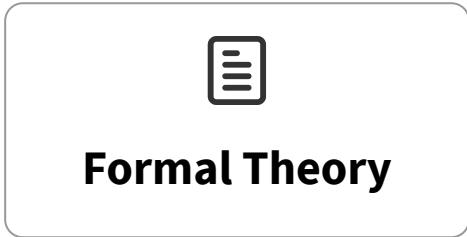
Contrived setting

A closed system that models the operation of the concrete system but without participants

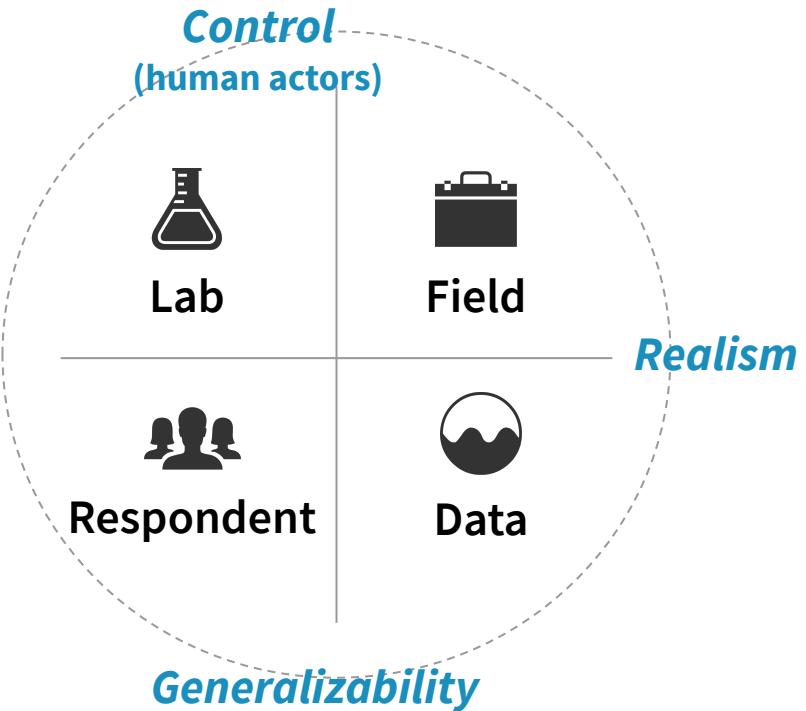
Behaviour must also be modeled, so all behavioural parameters must be known in advance – based on previous empirical evidence

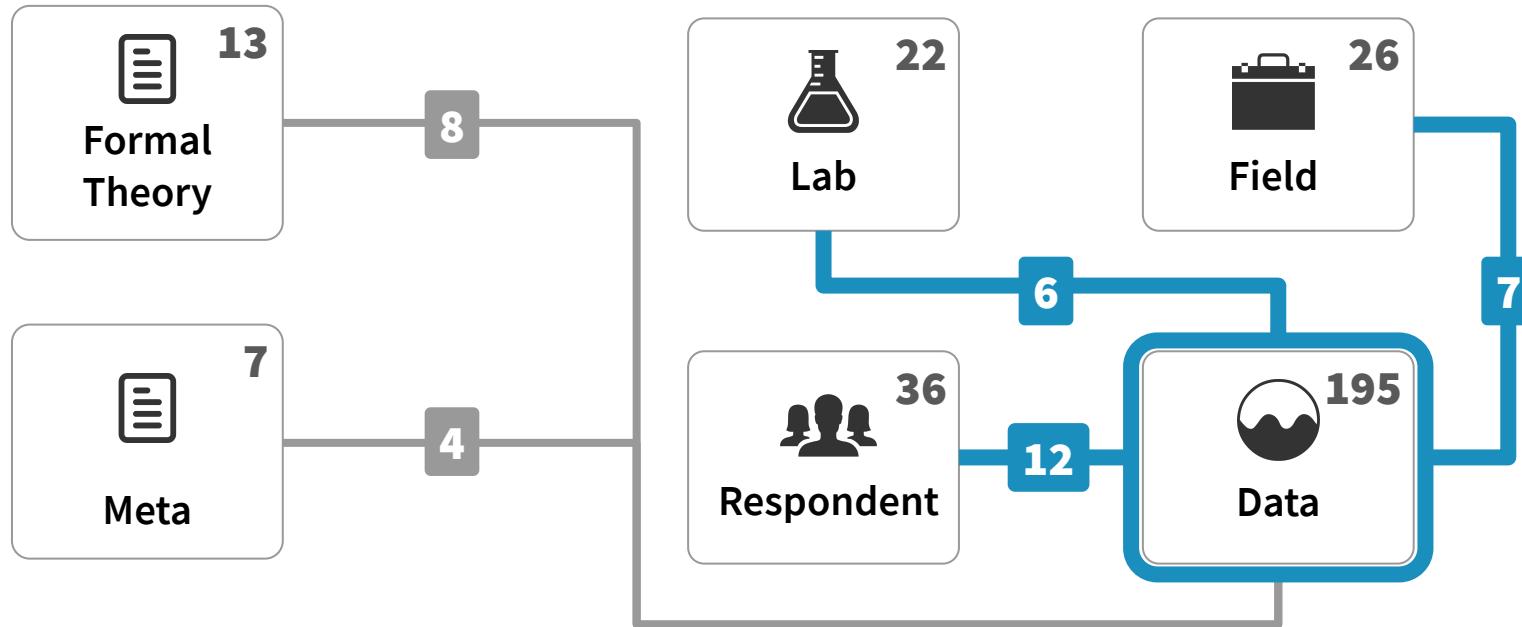


Non-Empirical



Empirical





48 | 37 Data Papers Used **Triangulation**

Triangulation: “The basic idea underpinning the concept of triangulation is that the phenomena under study can be understood best when approached with a variety or a combination of research methods. Triangulation is most commonly used in data collection and analysis techniques, but it also applies to sources of data. It can also be a rationale for multiple investigators in team research.”

<http://methods.sagepub.com/reference/sage-encyc-qualitative-research-methods/n468.xml>

But 17/60 authors that responded to our survey did not understand the concept “triangulation”

Sequential explanatory strategy: e.g., quantitative analysis of trace data followed by qualitative analysis of interview data (latter helps explain the former)

Sequential exploratory strategy: e.g., analysis of qualitative data from surveys followed by analysis of quantitative trace data (for testing emerging theory, explain early exploratory findings)

Concurrent triangulation strategy: different methods used concurrently, improve validity

Objectives/
Research Questions

DS view on SE

DS feasibility for SE

Knowledge/Theory

Review DS literature

Visual Abstract

SE understanding

SE contribution types

Methodology

- unit of analysis

SE papers

- data collection

Extract from papers

- data analysis

Identify DS elements

Validation

ICSE best papers

Extract from papers

Define VAs

Peer/Author review

Cluster papers

Peer review

Define Visual Abstract Template

Analyse ICSE best papers from DS perspective

51 | Research process map (Zagalsky, Runeson, Murphy, Storey)

Easterbrook, Singer, Storey & Damian, “Selecting empirical methods for software engineering research,” 2008

Research Design, by John W. Cresswell

Educational Research, by John W. Cresswell

Joseph. E. McGrath. Methodology matters: Doing research in the behavioral and social sciences. In Readings in Human-Computer Interaction: Toward the Year 2000, Ronald M. Baecker (Ed.). Morgan Kaufmann, 152–169.

Philip J. Runkel and Joseph E. McGrath: “Research on Human Behavior: A Systematic Guide to Method” (1972)

Stol, Ralph & Fitzgerald, “Grounded theory in software engineering research: A critical review and guidelines,” 2016

Runeson & Höst, “Guidelines for conducting and reporting case study research in software engineering,” 2008

Zelkowitz & Wallace, 1998 “Experimental Models for Validating Technology”

Shaw, 2002 “Writing good software engineering research papers”

Vessey, Ramesh and Glass 2005, “A unified classification system for research in the computing disciplines”

Smite, Wohlin, Gorschek, and Feldt, 2010 “Empirical evidence in global software engineering: a systematic review”

Wohlin and Aurum, 2015 “Towards a decision-making structure for selecting a research design in empirical software engineering”

Siegmund , Siegmund and Apel, 2015 “Views on internal and external validity in empirical software engineering”

Stol and Fitzgerald, 2018 “The ABC of Software Engineering Research”

Additional reading



Methods Activity: Categorizing Papers Using the Socio-Technical Research Strategy Framework

For the SBES 2019 Best Papers, which strategies do they use?

Which research methods do you primarily use in your studies?



Lab studies (artificial setting with human subjects)



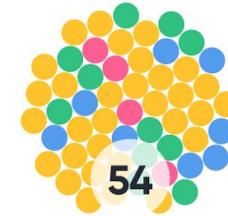
Data studies (no human subjects)



Field case studies or experiments (with human subjects)



Meta studies (theories or literature reviews)



Respondent studies (surveys, interviews)



Formal methods

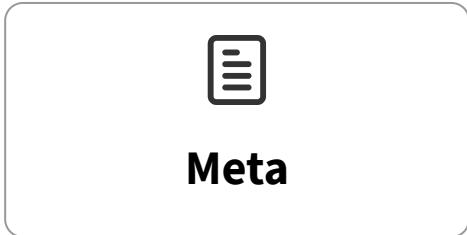
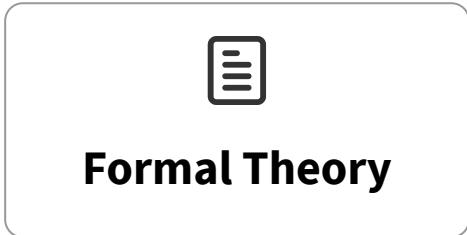
For your research, do you collaborate with industry ?

- Always
- Sometimes
- Never
- Unknown

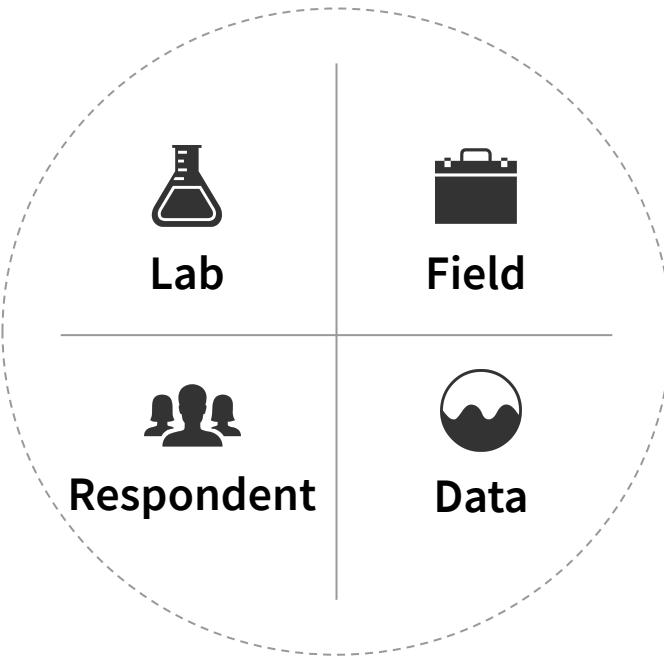
Screenshot

243

Non-Empirical

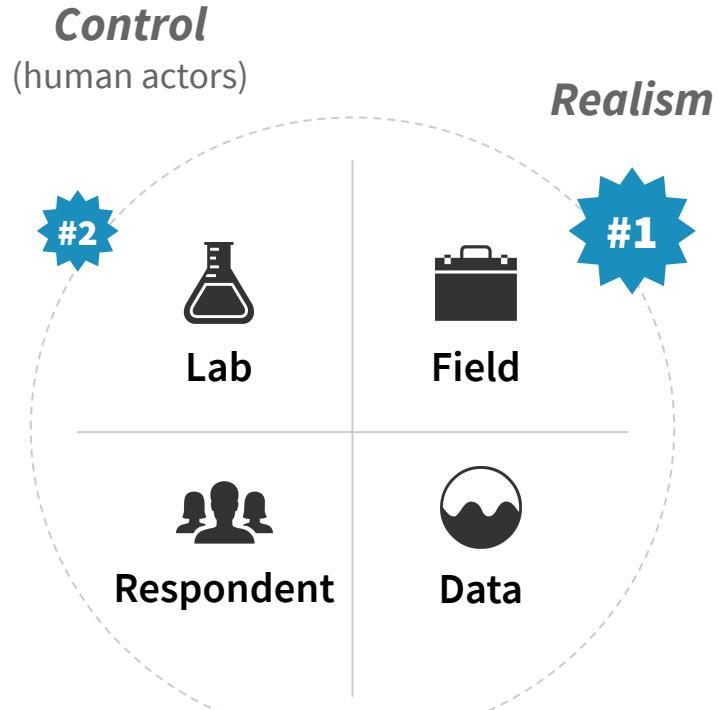


Empirical



56 | Write the title of the paper here

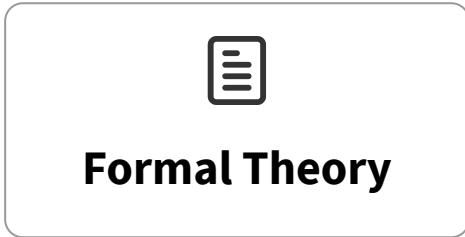
CloudBuild



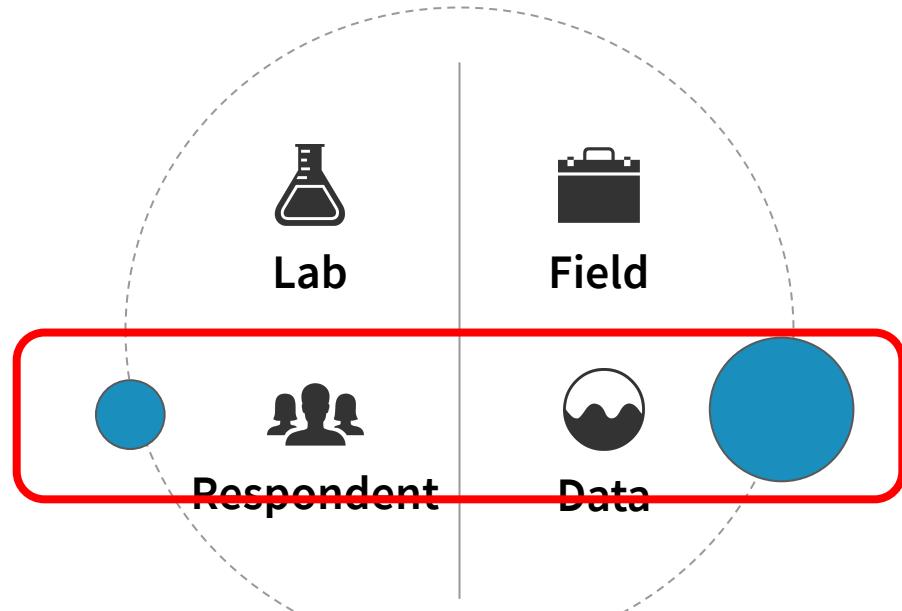
57 | The Methods We Chose

Lebeuf, Voyloshnikova, Herzig & Storey:
“Debugging, and Optimizing Distributed
Software Builds: A Design Study”, ICMSE 2018

Non-Empirical



Empirical



| Which primary and secondary strategies do the authors report
(use size to indicate if one study is more of a focus)?

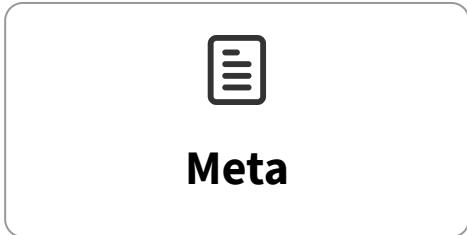
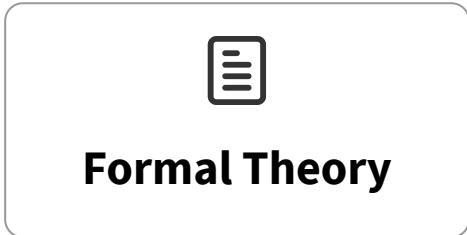
Sequential explanatory strategy: e.g., quantitative analysis of trace data followed by qualitative analysis of interview data (latter helps explain the former)

Sequential exploratory strategy: e.g., analysis of qualitative data from surveys followed by analysis of quantitative trace data (for testing emerging theory, explain early exploratory findings)

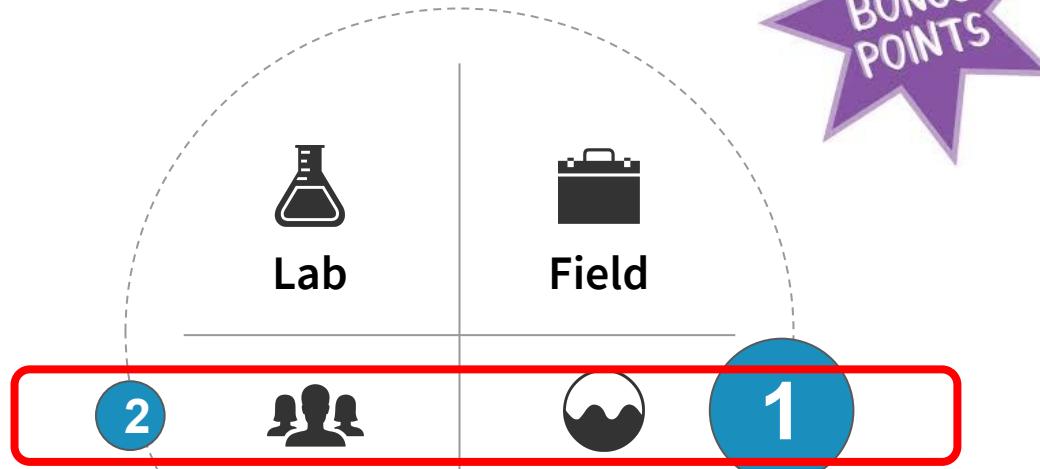
Concurrent triangulation strategy: different methods used concurrently, improve validity



Non-Empirical



Empirical



Sequential explanatory strategy!

60 | If mixed methods, which design strategy do they use?



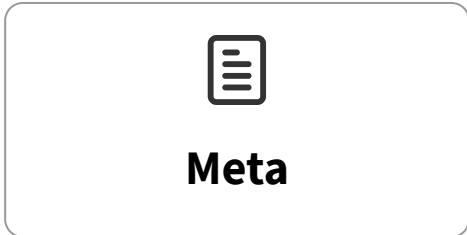
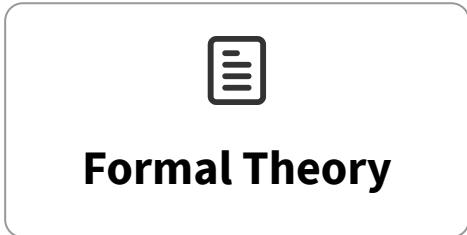
Of the
Data Only Papers...

How many authors
mention
Developers

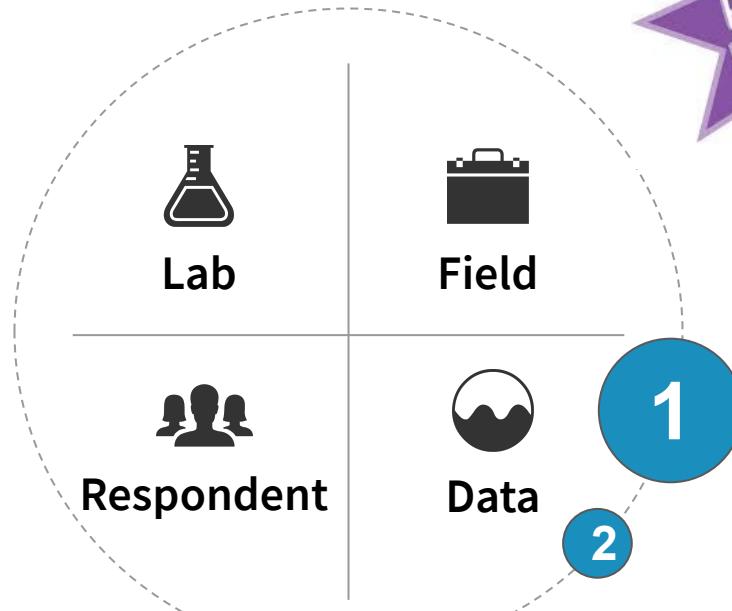
(But don't study human
research subjects)



Non-Empirical



Empirical



1

2

Yes, aim to improve dev productivity

62 | If data strategies only, do they mention human stakeholders?

Which **research quality criteria** did the authors achieve in their choice of strategies (i.e., generalizability, control, realism)?

Could they have used **different strategies** in their research design to answer their research questions?

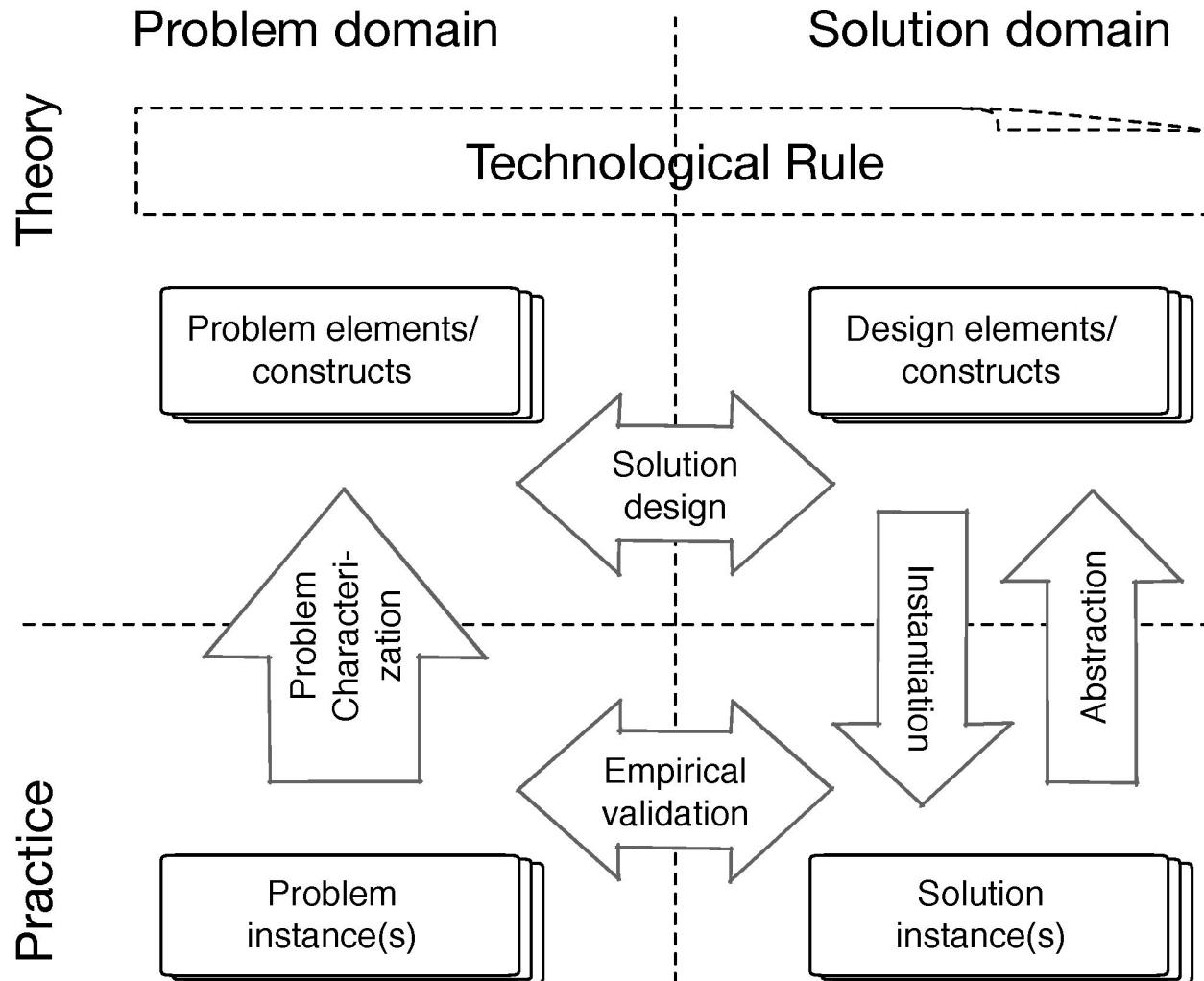
What else could we have asked about the choice of methods?

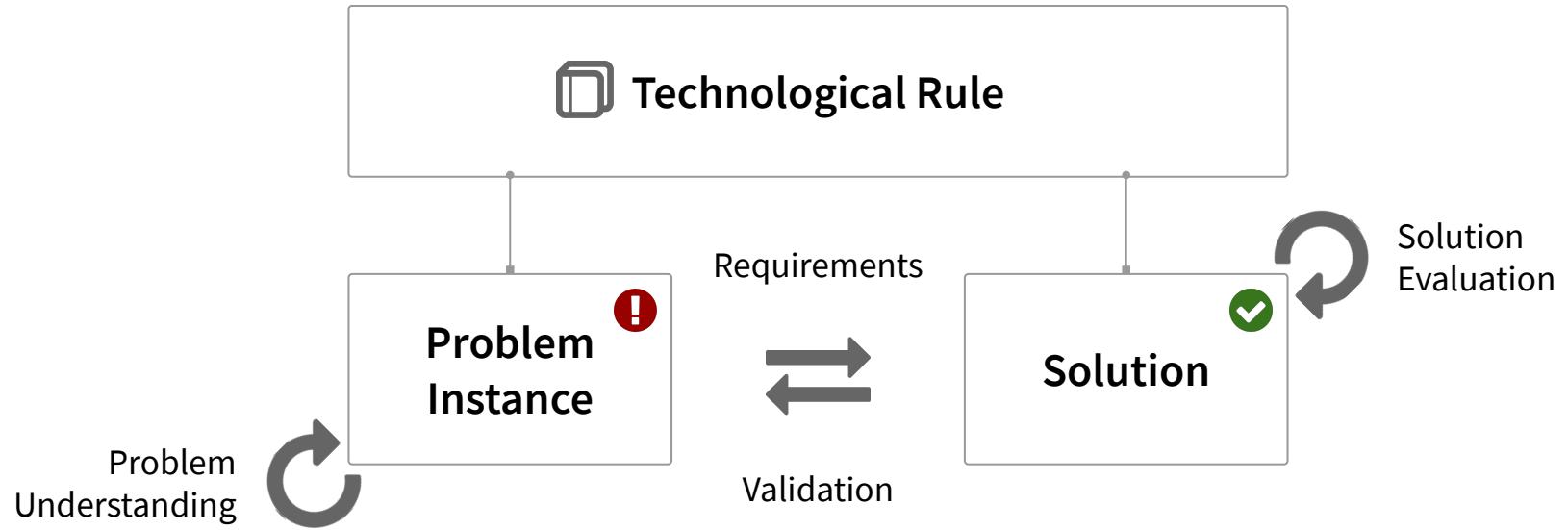
Motivation

Paradigms (activity)

Methods (activity)

 **Contributions (activity)**







Technological Rule

(Theory Fragment)

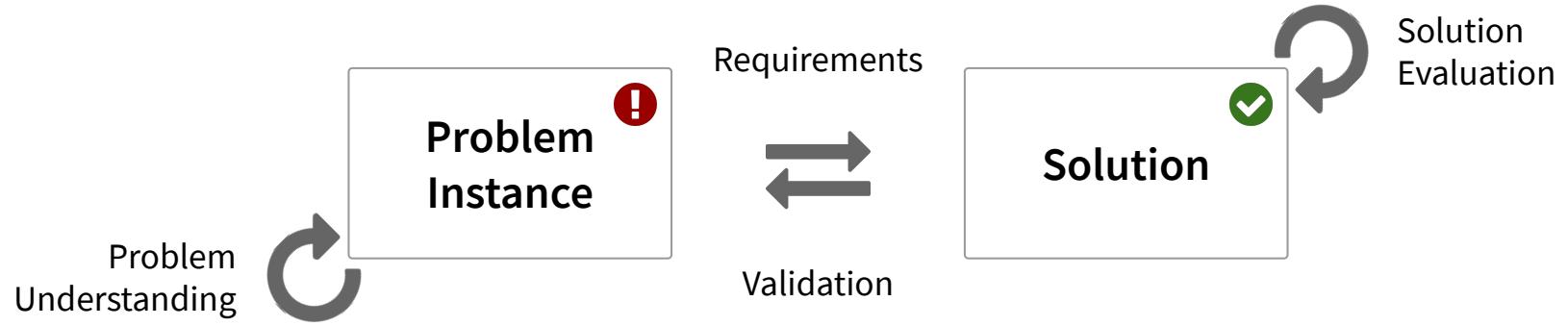
To reduce errors in open source projects
use continuous integration.

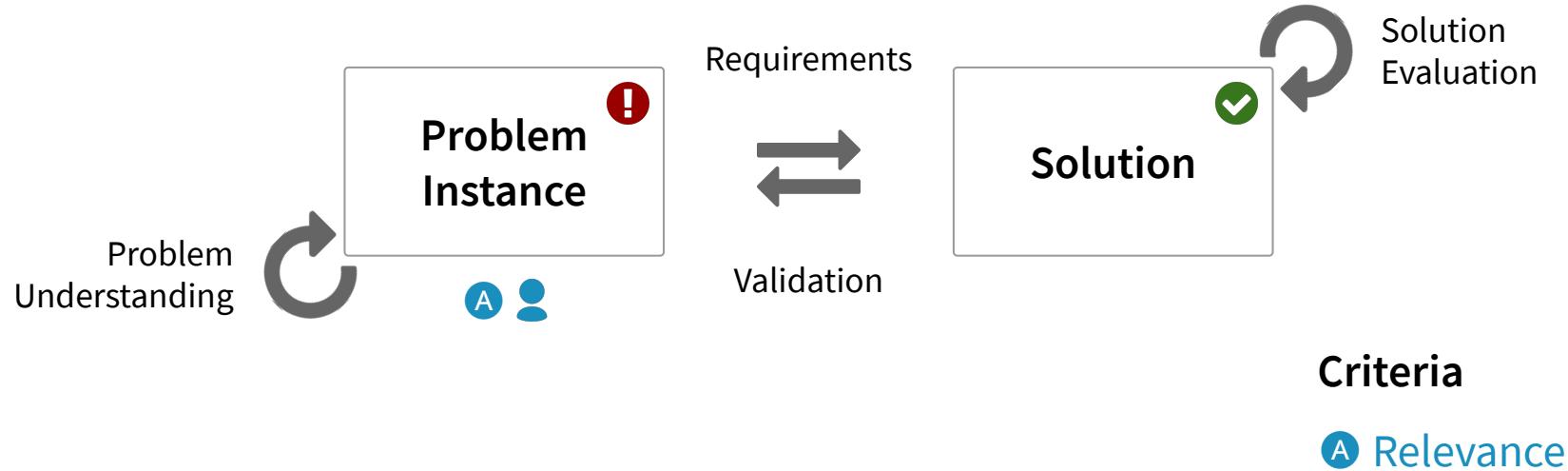
To achieve an effect in a given context use / do intervention.

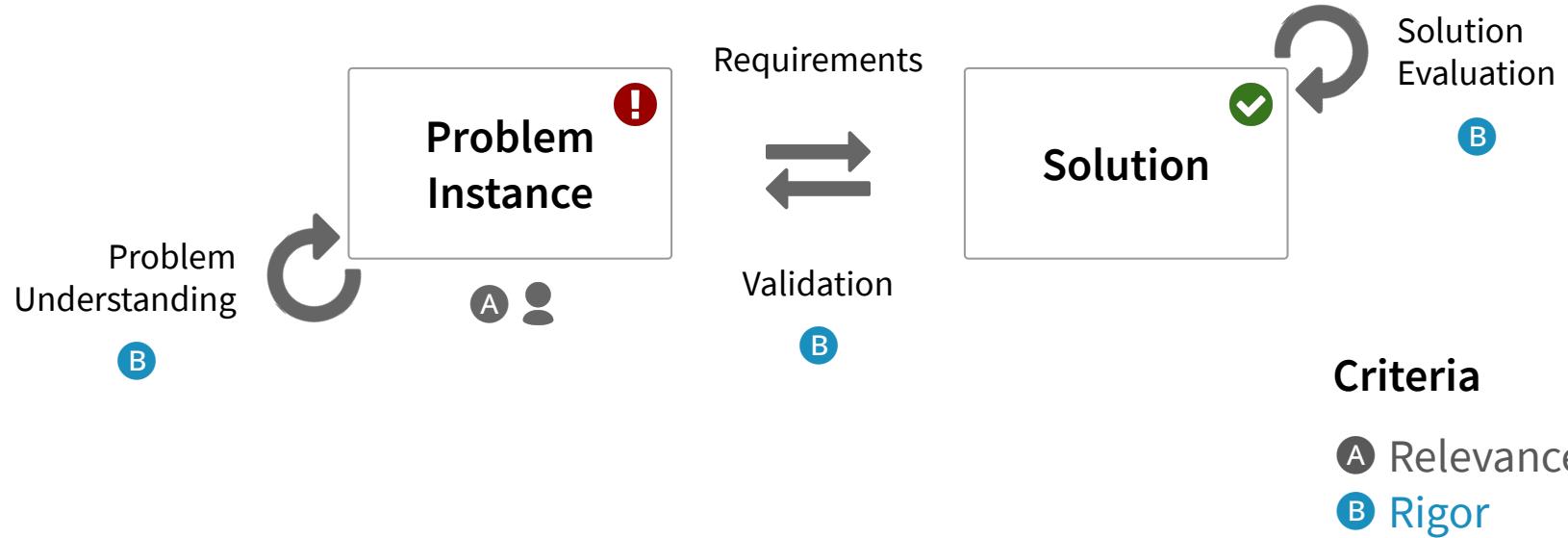
x

y

z







Construct validity

Internal validity

External validity

Reliability: would the study yield the same results if done by different researchers?

Triangulation

Prolonged contact with participants

Member checking

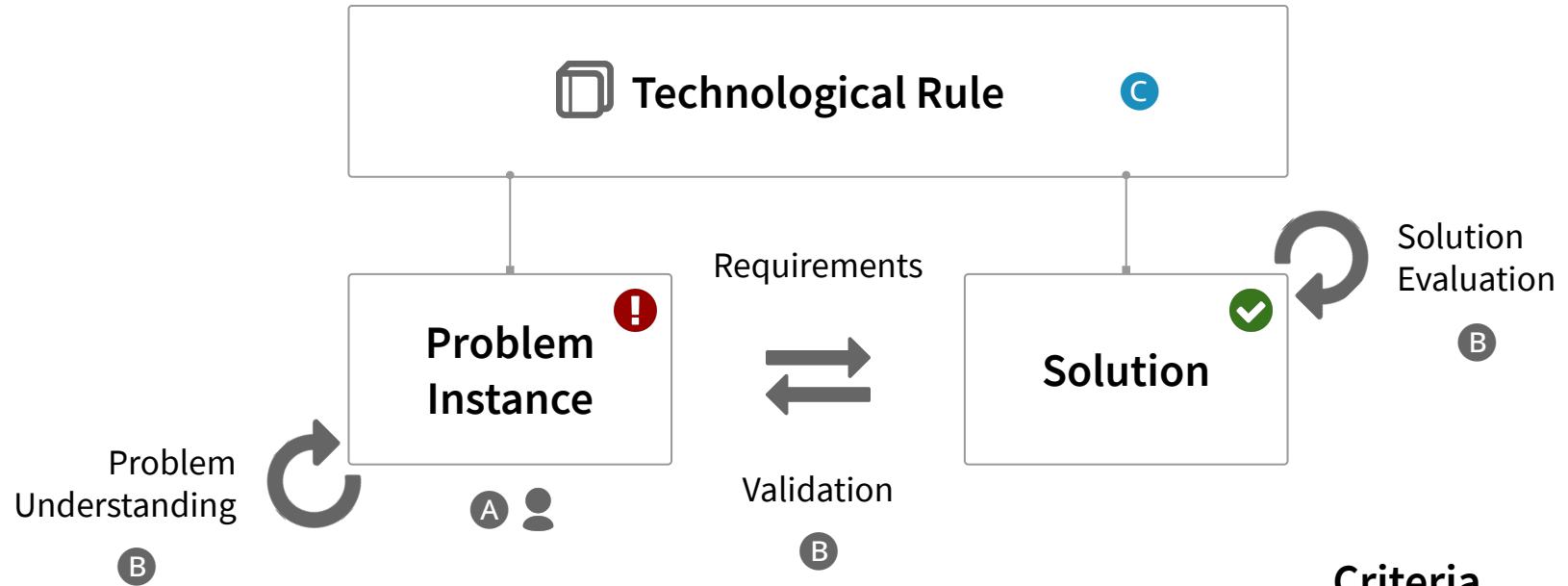
Peer debriefing (plan ahead for this!)

Rich, thick descriptions

External auditor (also need to plan)

Clarify bias (report researcher bias)

Report discrepant information



Criteria

- (A) Relevance
- (B) Rigor
- (C) Novelty

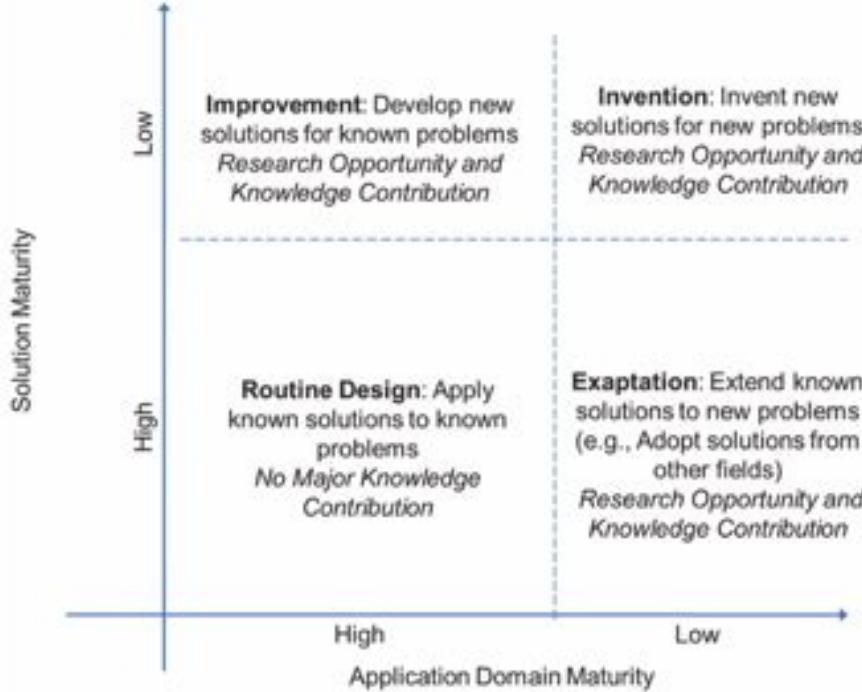


Figure 3. DSR Knowledge Contribution Framework

74 | Categorizing design science knowledge contribution

Refinement of an existing rule? (context, intervention)

Confirmation of an existing rule?

Refuting an existing rule?

New rule?

COMPONENTS OF AN EFFECTIVE VISUAL ABSTRACT

Summarize Key Question
Being Addressed

Impact of treating Iron Deficiency Anemia Before Major Abdominal Surgery

Decreased Need for
Blood Transfusions



31% → 12%
(percent of patients)

Shorter Hospital
Length of Stay



9.7 → 7.0
(days)

Recovery of Hemoglobin
(Hb) post-discharge



+0.9 → +1.9
(Hb change at 4 weeks)

Summary of
Outcomes

State Outcome Comparison

Visual Display of Outcome

Data of Outcome (Units)

Who Created the Visual
Abstract (often the journal)

Author, Citation

Froessler et al. Ann Surg. July 2016

ANNALS OF SURGERY
A Monthly Review of Surgical Science Since 1883

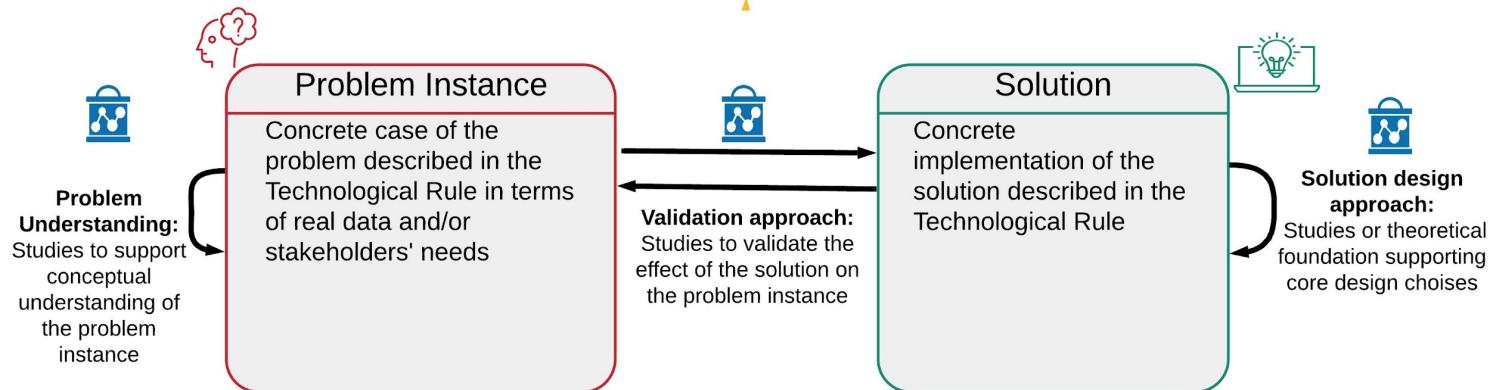
@AnnalsofSurgery | Copyright © 2016 Wolters Kluwer Health, Inc. All rights reserved. Published by Lippincott Williams & Wilkins, Inc.

76 | Visual Abstract Concept from Medicine

<https://www.surgeryredesign.com/resources/>



Technological rule: To achieve effect/change in situation/context apply solution/intervention



Relevance: Characteristics of the context that are likely to impact applicability and potential value of the proposed solution



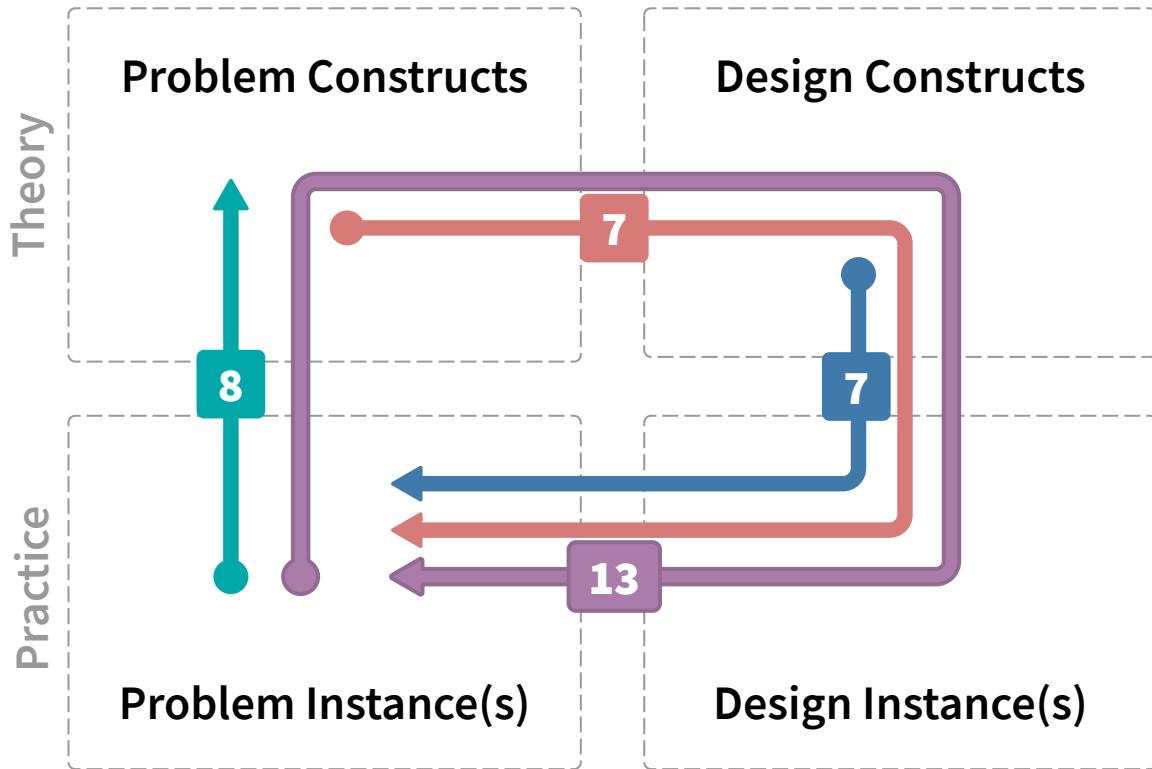
Rigor: Characteristics of the three knowledge creating activities (problem understanding, solution design and in context evaluation) that adds to the strength of the empirical support of the Technological Rule



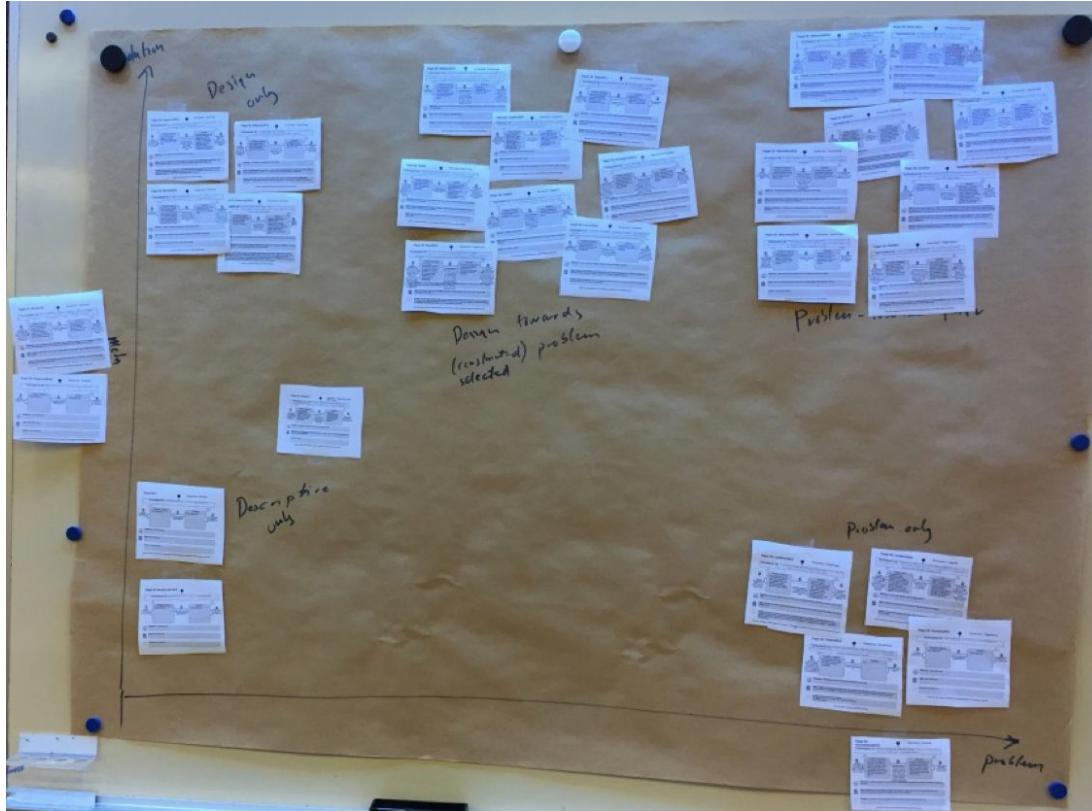
Novelty: Positioning of the Technological Rule in terms of previous knowledge



- Problem Solution
- Solution Validation
- Solution Design
- Descriptive
- Meta

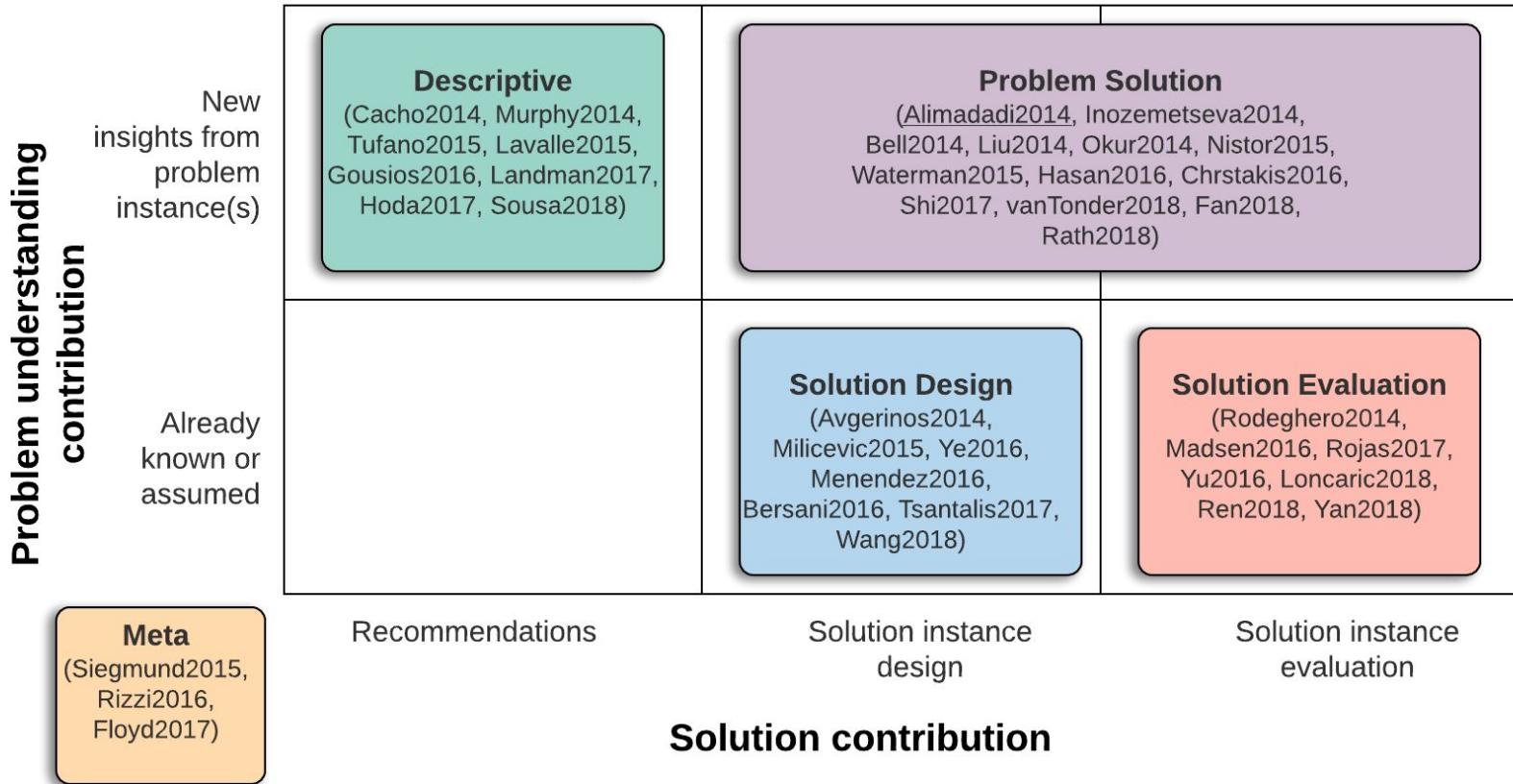


78 | How were papers clustered?



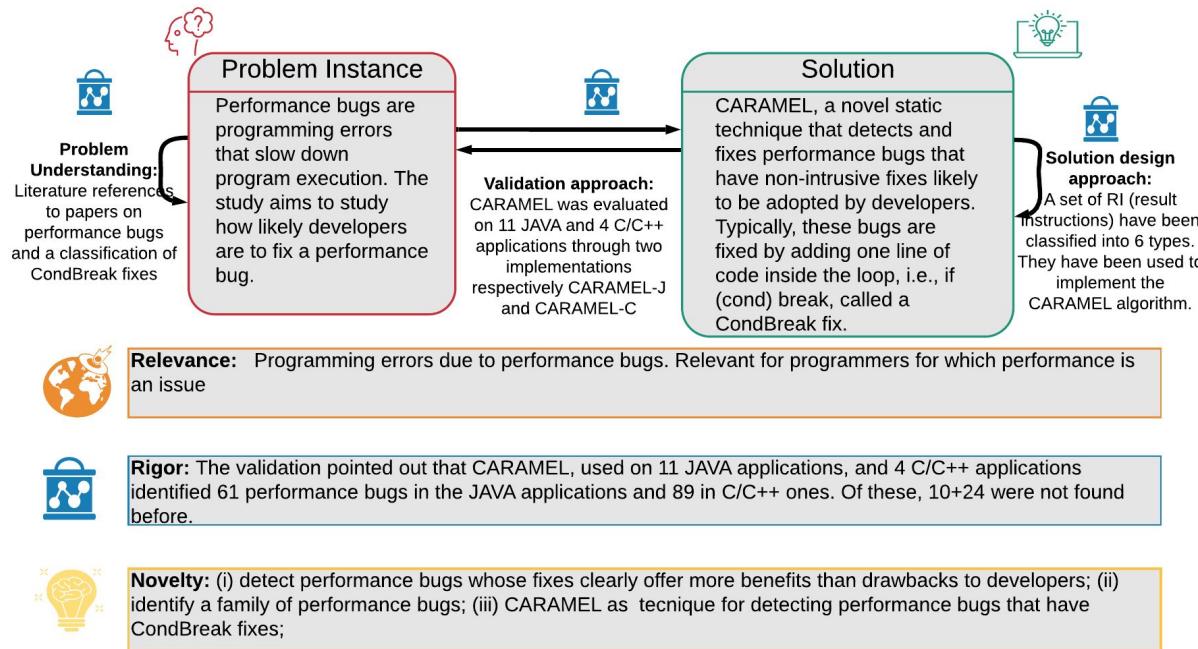
79 | How we clustered

Clustered Papers by Design Science Contributions





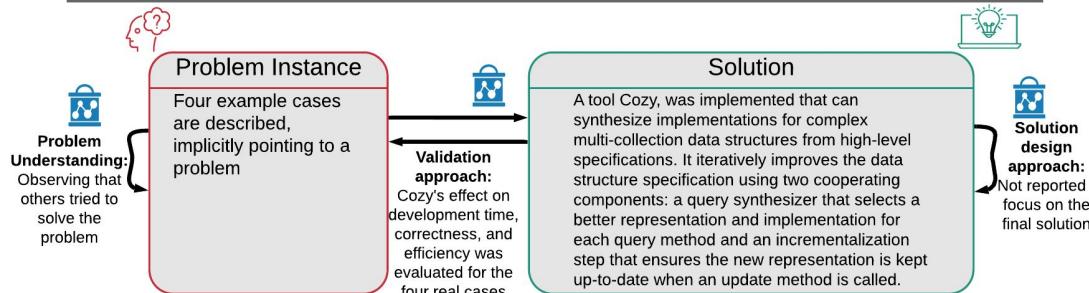
Technological rule: To fix performance bug problems caused by unnecessary loop execution use CARAMEL, which adds a break within loops



Paper Title: CARAMEL: Detecting and Fixing Performance Problems That Have Non-Intrusive Fixes



Technological rule: To synthesize data structures that track subsets and aggregations of multiple related collections alternate steps of query synthesis and incrementalization



Relevance: Data structure problems, especially in domains like user interfaces or web services where software must manage some internal state and also handle asynchronous events.



Rigor: Proof of concept demonstrated in four real cases

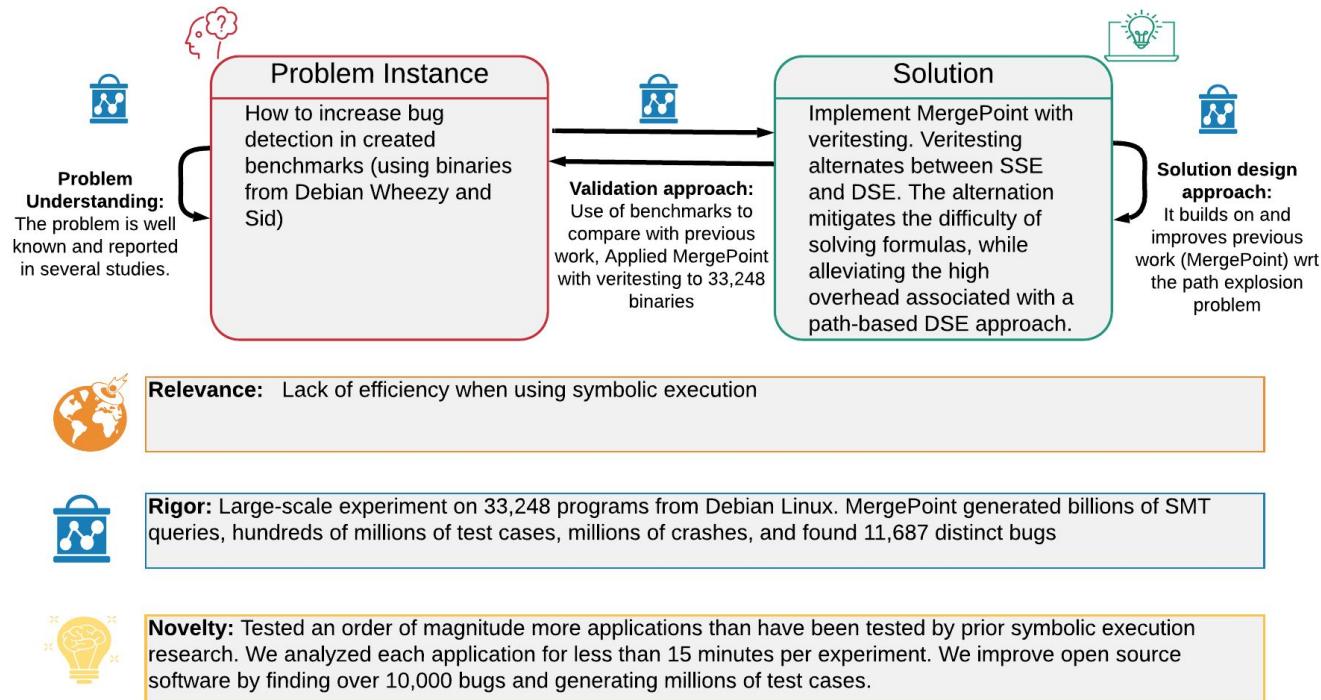


Novelty: It is a new technique for data structure synthesis that overcomes many of the limitations of previous work

Paper Title: Generalized Data Structure Synthesis

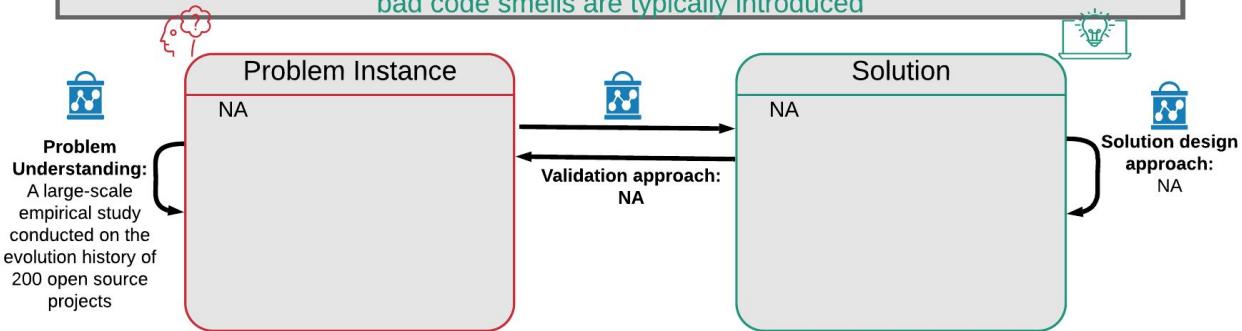


Technological rule: To increase efficiency and effectiveness when using symbolic execution alternate between static and dynamic symbolic execution





Technological rule: To better plan activities for improving design and source code quality when developing software utilise the knowledge gained in this study about when and why bad code smells are typically introduced



Relevance: The study context is the change history of 200 projects belonging to three software ecosystems, namely Android, Apache, and Eclipse



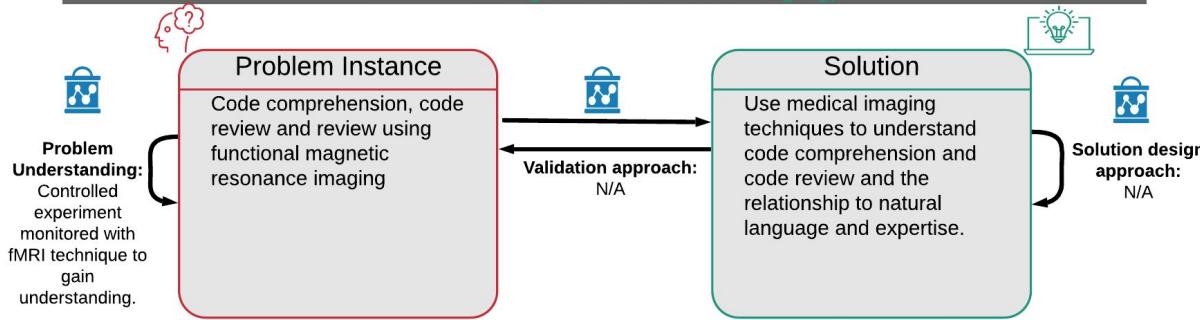
Rigor: Validity is ensured because a large set of 200 projects concerning the analysis of code smells and of their evolution has been investigated. Projects were extracted from three ecosystems: android, apache and eclips.



Novelty: First comprehensive empirical investigation into when and why code smells are introduced in software projects.



Technological rule: To understand how human brain processes software engineering tasks when making subjective human judgments use medical imaging techniques (fMRI - functional magnetic resonance imaging)



Relevance: Code review and comprehension vs prose review activities



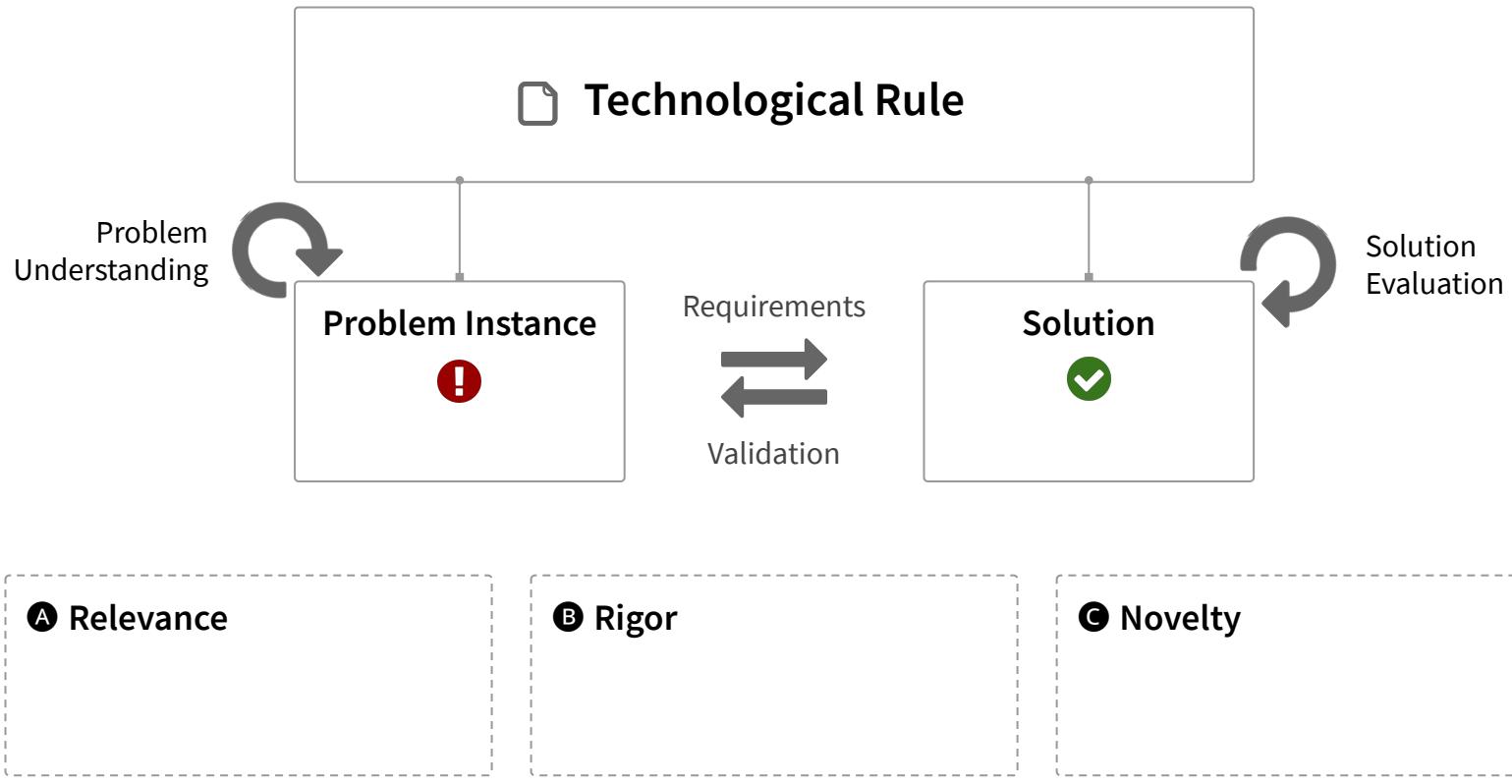
Rigor: In a controlled experiment involving 29 participants, authors examine code comprehension, code review and prose review using functional magnetic resonance imaging.



Novelty: Measure brain activity for carrying out different tasks

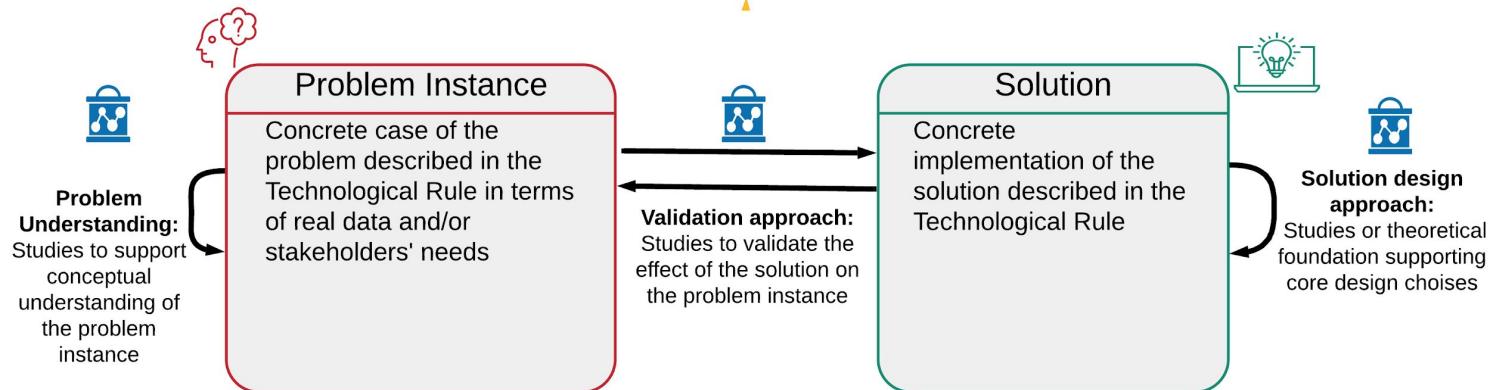


Activity: Applying the Design Science Lens





Technological rule: To achieve effect/change in situation/context apply solution/intervention



Relevance: Characteristics of the context that are likely to impact applicability and potential value of the proposed solution

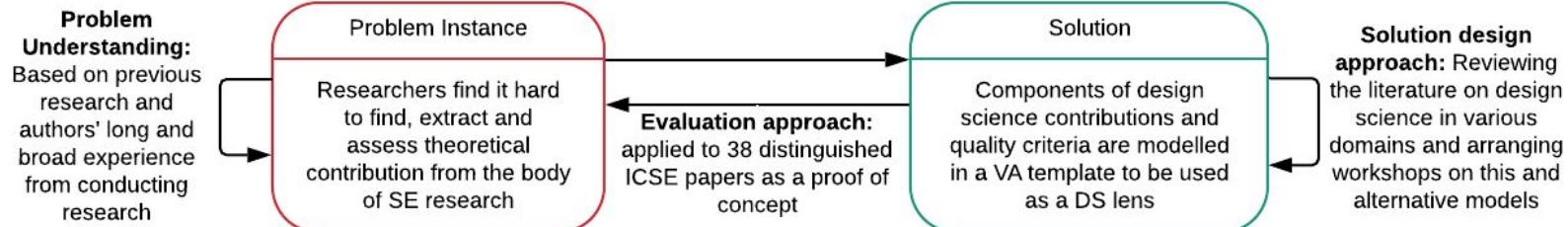


Rigor: Characteristics of the three knowledge creating activities (problem understanding, solution design and in context evaluation) that adds to the strength of the empirical support of the Technological Rule



Novelty: Positioning of the Technological Rule in terms of previous knowledge

Technological rule: To better assess and communicate research contributions within software engineering apply a design science lens



Relevance: Researchers aiming at conducting and disseminating industry-relevant SE research

Rigor: The problem is well known and the proposal is based on thorough review of design science litterature, no validation of the effects on communication and quality assessment was made

Novelty: The conceptualization of design science for software engineering and the VA model is novel, although it builds on previous discussions on the topic (Wieringa, Hevner, vanAken)

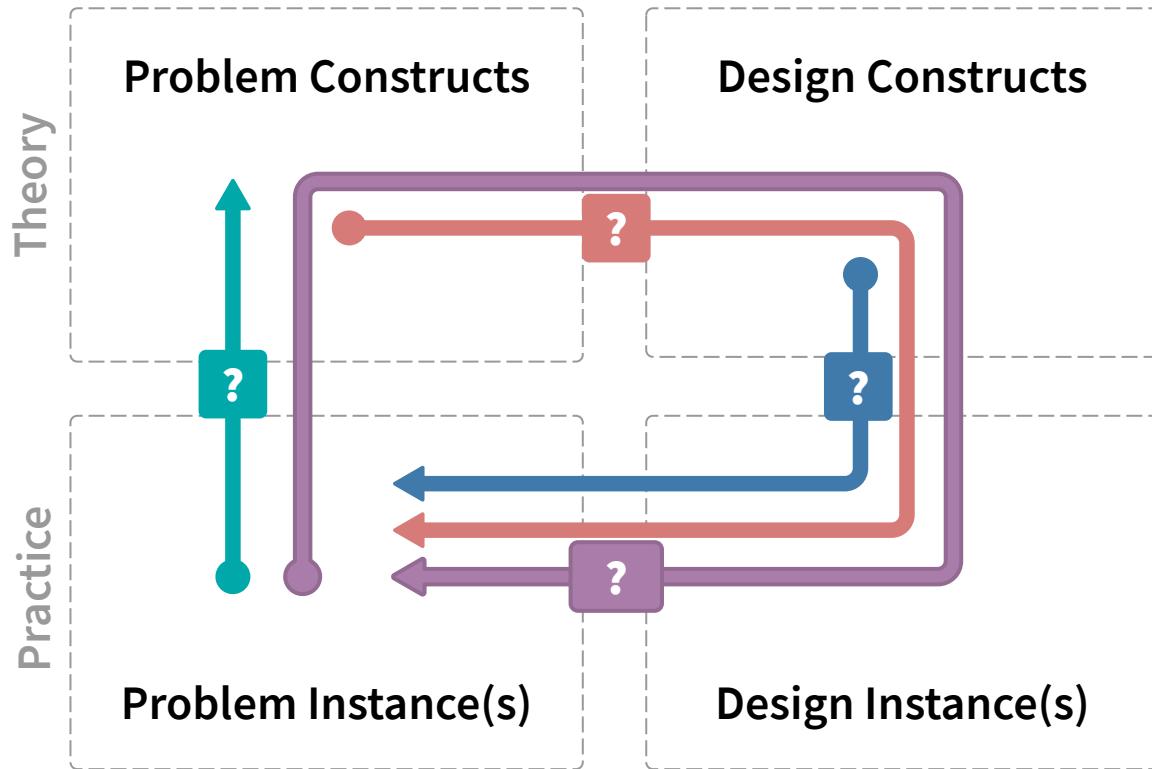


5

SBES Best Papers



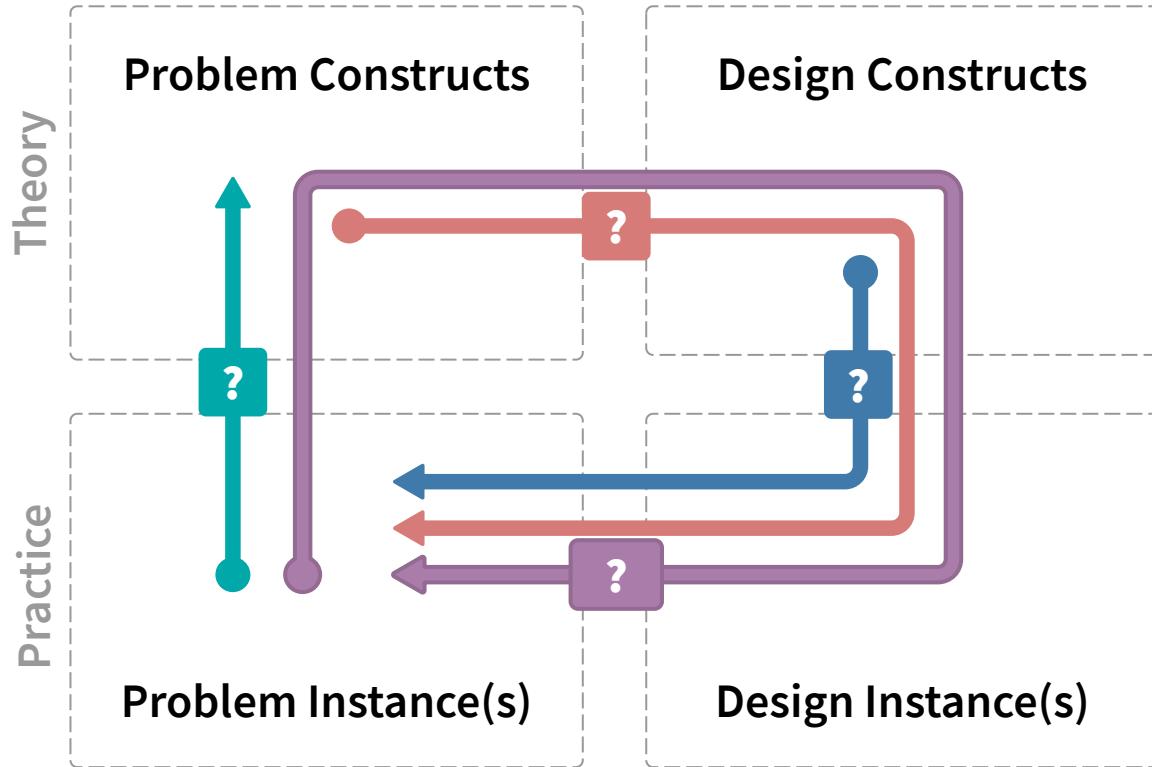
- Problem Solution
- Solution Validation
- Solution Design
- Descriptive
- Meta



91 | Which contribution type is produced by this research?

Consider Stakeholders

- Problem Solution
- Solution Validation
- Solution Design
- Descriptive
- Meta



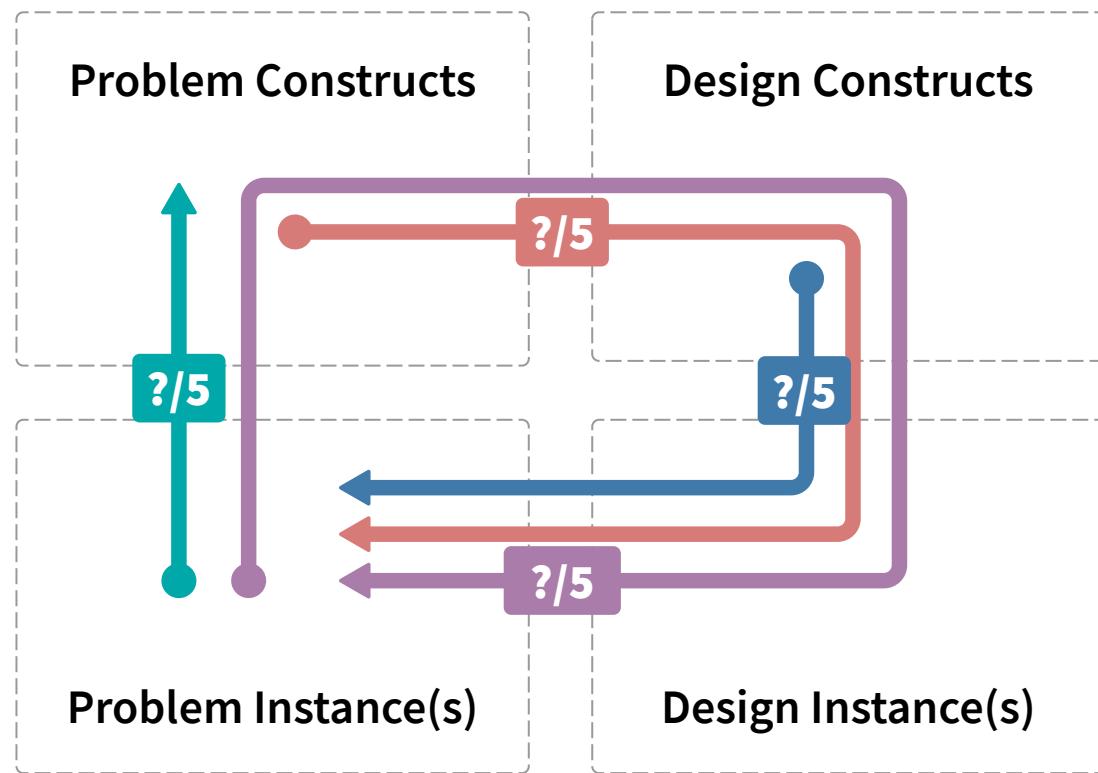
92 | Relevance to stakeholders?



?/5

Consider Stakeholders

- Problem Solution
- Solution Validation
- Solution Design
- Descriptive



What **challenges** did you face filling out the visual abstract?

How does it compare to other types of **structured abstracts**?

Do you think this would be useful in a **systematic literature review** for linking hierarchically technological rules?

A structured abstract is an abstract with distinct, labeled sections (e.g., Introduction, Methods, Results, Discussion) for rapid comprehension.

In comparison, visual abstracts provide additional insights with a more flexible reading order

PubMed 24792780[uid]

RSS Save search Advanced

Display Settings: Abstract Send to:

Clin Toxicol (Phila). 2014 Jun;52(5):525-30. doi: 10.3109/1556368502014_913175. Epub 2014 May 5.

Evaluation of dexmedetomidine therapy for sedation in patients with toxicological events at an academic medical center.

Mohorn PL¹, Vakkalanka JP, Bushton W, Hardison L, Woloszyn A, Holstein C, Corbett SM.

Author information

Abstract

INTRODUCTION: Although clinical use of dexmedetomidine (DEX), an alpha2-adrenergic receptor agonist, has increased, its role in patients admitted to intensive care units secondary to toxicological sequelae has not been well established.

OBJECTIVES: The primary objective of this study was to describe clinical and adverse effects observed in poisoned patients receiving DEX for sedation.

METHODS: This was an observational case series with retrospective chart review of poisoned patients who received DEX for sedation at an academic medical center. The primary endpoint was incidence of adverse effects of DEX therapy including bradycardia, hypotension, seizures, and arrhythmias. For comparison, vital signs were collected hourly for the 5 h preceding the DEX therapy and every hour during DEX therapy until the therapy ended. Additional endpoints included therapy duration, time within target Richmond Agitation Sedation Score (RASS); and concomitant sedation, analgesia, and vasopressor requirements.

RESULTS: Twenty-two patients were included. Median initial and median DEX infusion rates were similar to the commonly used rates for sedation. Median heart rate was lower during the therapy (82 vs. 93 beats/minute, p < 0.05). Median systolic blood pressure before and during therapy was similar (111 vs. 109 mmHg, p = 0.745). Five patients experienced an adverse effect per study definitions during therapy. No additional adverse effects were noted. Median time within target RASS and duration of therapy was 6.5 and 44.5 h, respectively. Seventeen patients (77%) had concomitant use of other sedation and/or analgesia with four (23%) of these patients requiring additional agents after DEX initiation. Seven patients (32%) had concomitant vasopressor support with four (57%) of these patients requiring vasopressor support after DEX initiation.

CONCLUSION: Common adverse effects of DEX were noted in this study. The requirement for vasopressor support during therapy warrants further investigation into the safety of DEX in poisoned patients. Larger, comparative studies need to be performed before the use of DEX can be routinely recommended in poisoned patients.

PMID: 24792780 [PubMed - indexed for MEDLINE]



Hierarchy of Technological Rules (CI)

Problems

Awareness

Improve testing

Environment
management



Solutions

Traceability

Branch coverage

Provisioning

Feldt & Magazinius, "Validity Threats in Empirical Software Engineering Research-An Initial Survey," 2010

Siegmund, Siegmund & Apel, "Views on internal and external validity in empirical software engineering," 2015

Shneiderman, "Twin-Win Model: A human-centered approach to research success," 2018

Sjøberg et al., "Building theories in software engineering," 2008

Stol & Fitzgerald, "Uncovering theories in software engineering," 2013

Ralph, "Possible core theories for software engineering," 2013

Hevner, "A three cycle view of design science research," 2007

Van Aken, "Management Research Based on the Paradigm of the Design Sciences: The Quest for Field-Tested and Grounded Technological Rules," 2004

“Using a visual abstract as a lens for communicating and promoting design science research in software engineering”, Storey, Engström, Höst, Runeson, Bjarnason, ESEM 2017. <http://chisel.cs.uvic.ca/pubs/storey-ESEM2017.pdf>

“A review of software engineering research from a design science Perspective”, Engström, Storey, Runeson, Höst, Baldassarre, Arxiv 2019. <http://arxiv.org/abs/1904.12742>

“Methodology Matters: How We Study Socio-Technical Aspects in Software Engineering.”, Margaret-Anne Storey, Courtney Williams,, Neil A. Ernst, Alexey Zagalsky and Eirini Kalliamvakou. 2019. Arxiv. <https://arxiv.org/abs/1905.12841>

Twitter: @margaretstorey

<https://github.com/margaretstorey/cbsoft2019tutorial/>

| Concluding remarks



“A paradigm is a shared world view that represents the beliefs and values in a discipline and that guides how problems are solved.”

– Schwandt, 2001



Scientific method

Evidence-based reality

Theory verification and falsification

Quantitative over qualitative

Reality is subjective and experiential

Theory generation

Biases are expected and made explicit

Qualitative over quantitative





Change oriented

Collaborative

Shaped by political and social lenses

Qualitative and quantitative

Pragmatism

Problem centered • Real-world practice oriented

