Research Methods in Software Engineering

An Introduction into Philosophy of Science for Software Engineers

Based in parts on material from a joint work with:
Antonio Vetrò (Nexa Center for Internet and Society, Politecnico di Torino)
Andreas Jedlitschka (Fraunhofer Institute for Experimental Software Engineering)
Natalia Juristo (Politecnic University of Madrid)

Daniel Méndez

@mendezfe

www.mendezfe.org

Based on slides by @SMEasterbrook and @ethanwhite

Ground rule

Whenever you have questions / remarks,
don’t ask Google Deutschland, but
share them with the whole group.

Goal of the (invited) lectures

What you have learnt so far

- Methods for empirical software engineering
- Theory building

Focus:
What & How

Get a "bigger picture" by better understanding

- Fundamental principles, concepts, and terms in philosophy of science
- The (historical) context of research strategies
- Broader perspective on empirical Software Engineering

Focus:
Why

Exemplary, more philosophical questions

- What is *truth*? Is there such a thing as universal/absolute truth? (i.e. assuming that there is a physical reality which represents “truth”, are we able to completely capture it via theories?)
- How can we achieve *scientific progress*?
- Which *research methods* should we apply?
- What is a *suitable (empirical) basis*?
- When is an observation *objective?* Is there really objectivity?
- How much *relevance/impact* can we achieve? What does relevance mean?
- What trade-offs do I need to make when designing a study?
- …

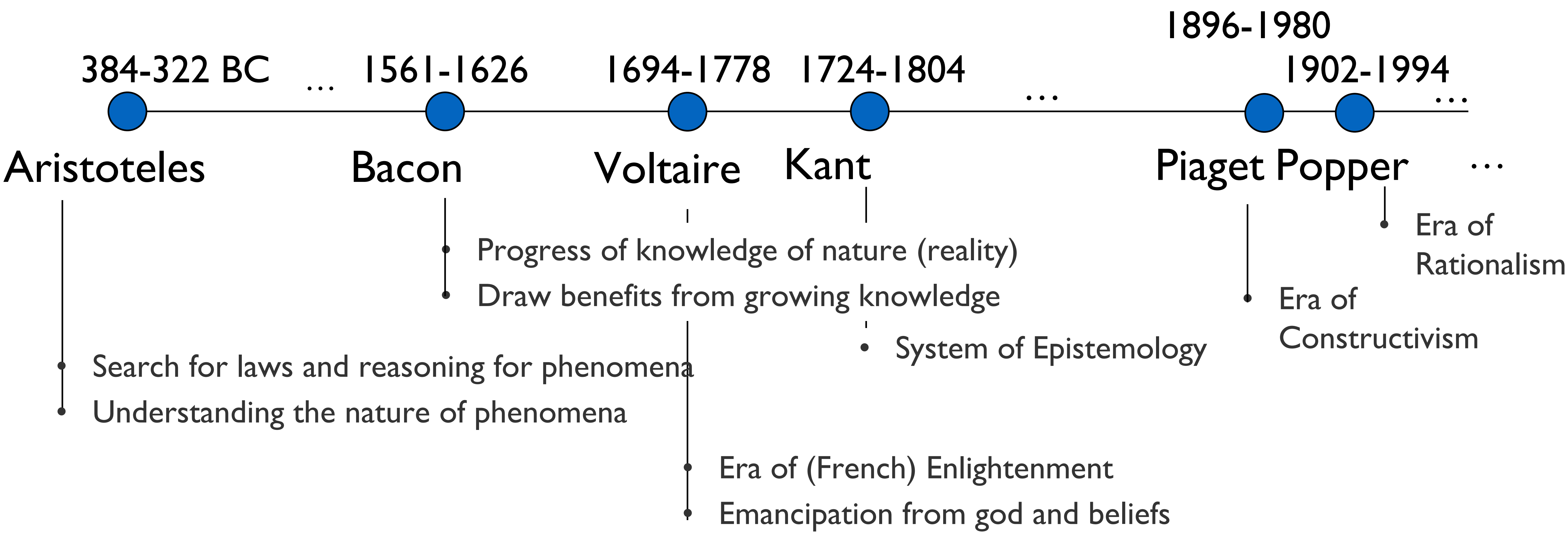
Outline

- Science (in a Nutshell)
- Philosophy of Science - a Historical Perspective
- Key Take Aways
- From Philosophy of Science to Empirical Software Engineering
- Empirical Software Engineering Processes
- Current Challenges in Empirical Software Engineering

Outline

- **Science (in a Nutshell)**
- Philosophy of Science - a Historical Perspective
- Key Take Aways
- From Philosophy of Science to Empirical Software Engineering
- Empirical Software Engineering Processes
- Current Challenges in Empirical Software Engineering

“Science” wasn’t built in a day…



- Science is a human undertaking for the search of knowledge (by portraying reality and its laws)
- It needs to be considered in a historical context

  Knowledge growth

  Increased understanding of scientific working (and what science eventually is)

Stress-fields in science

Ontology	Epistemology	Ethics
Questions on the "*being*"	Questions on *knowledge* and the *"scientific discovery"*	Questions on *actions and morality*

From: Orkunoglu, 2010

Stress-fields in science

Ontology	Epistemology	Ethics
Is there a world independent of subjectivity?	From where do discoveries result? From experiences?	From where does ethics result? Does there exist something like universal ethics?
Realism	Rationalism	Normative Ethics
Idealism	Empiricism	Descriptive Ethics

From: Orkunoglu, 2010

Setting

Philosophy of science

Principle ways of working

Methods and strategies

Fundamental theories

Example

Epistemology

↓

Empiricism

↓

Controlled Experimentation

↓

Hypothesis testing

↓

Statistics

Setting: Philosophy of science

Philosophy of science

Principle ways of working

Methods and strategies

Fundamental theories

Branch of philosophy concerned with

- foundations,
- methods, and
- implications

of/in science(s).

Central questions:

- What qualifies as scientific working?
- When are scientific theories reliable?
- What is the purpose of science?

Setting: Empirical Software Engineering

- Philosophy of science
- Principle ways of working
- Methods and strategies
- Fundamental theories

Setting: Empirical Software Engineering

Philosophy of science

Principle ways of working

Theory building and evaluation

… are supported by…

Methods and strategies

Methods and strategies

Fundamental theories

Analogy: Theoretical and Experimental Physics

Goals of the lecture

Get a basic understanding of

- philosophy of science
- implications for our discipline
- context of research methods and strategies

Philosophy of science

Principle ways of working

Methods and strategies

Fundamental theories

Theory building and evaluation

… are supported by…

Methods and strategies

Analogy: Theoretical and Experimental Physics

What is Science?

What do you think?

What is science about?

Systematically and objectively gaining, documenting/preserving, and disseminating knowledge

What is science about?

Systematically and objectively gaining, documenting/preserving, and disseminating knowledge

- Gaining knowledge by the systematic application of research methods
  - Reasoning by argument / logical inference
  - Empiricism (case studies, experiments,…)
  - …
- Research should:
  - Have a high scientific and / or practical relevance and impact
  - Be rigorous and correct

However…

- There is no universal way of scientific working (see *Pragmatism / epist. anarchy)* Method appropriateness depends on many non-trivial factors

What is science about?

Systematically and **objectively** gaining, documenting/preserving, and disseminating knowledge

In principle, we try to be objective (independent of subjective judgment)

However…

- There is nothing absolute about knowledge/“truth” (see *Scientific Realism*)
- Accepting documented knowledge depends on acceptance by (subjective) peers, often judging by desire for “novelty”, “aesthetics”, etc. (see *Post-Positivism*)

Accepting scientific results is also a social process

What is science about?

Systematically and objectively gaining, **documenting/preserving, and disseminating** knowledge

- Scientific knowledge needs to be disseminated
  - documented in a reproducible way following (often unwritten) rules,
  - evaluated (by peers), and
  - disseminated / communicated to the public

However…

- Science (and scientific publishing) is also part of an **economic system**

Can't Disrupt This: Elsevier and the 25.2 Billion Dollar A Year Academic Publishing Business

Twenty years ago (December 18, 1995), *Forbes* predicted academic publisher Elsevier's relevancy and life in the digital age to be short lived. In an article

Source: https://medium.com/@jasonschmitt/can-t-disrupt-this-elsevier-and-the-25-2-billion-dollar-a-year-academic-publishing-business-aa3b9618d40a#.il1kxilt9

In the end, science is a human undertaking

the scholarly kitchen

SEARCH

Home About

WHAT'S HOT AND COOKING IN SCHOLARLY PUBLISHING

AUTHORS, METRICS AND ANALYTICS, PEER REVIEW, RESEARCH, SOCIOLOGY

Is Peer Review a Coin Toss?

POSTED BY TIM VINES • DEC 8, 2011 • 52 COMMENTS

FILED UNDER ACADEMIC PUBLISHING, IMPACT FACTOR, PEER REVIEW, RESEARCH

As a managing editor, one of the most common questions I get is about the journal's acceptance rate. I'm typically puzzled by this because acceptance rate tells you very little about the likely fate of any one submission.

If all the submissions for a month went into a hat and a blindfolded editor pulled out a proportion of them to publish, the acceptance rate would indeed be a good indicator of an individual paper's chances of publication. In reality, if a paper is below the quality threshold for the journal, it's almost certain to be rejected; and if it's above that threshold, then it's almost certain to be accepted.

The interest in acceptance rate seems to be linked to the attitude that peer review is a coin toss, and hence the overall acceptance rate can predict the fate of each paper. Where does this attitude come from? Does it have any basis in reality?

ADVANCES IN INFORMATION SCIENCE

Bias in Peer Review

Carole J. Lee
Department of Philosophy, University of Washington, 361 Savery Hall, Seattle, WA 98195. E-mail: c3@uw.edu

Cassidy R. Sugimoto, Guo Zhang, and Blaise Cronin
School of Library and Information Science, Indiana University, 1320 East Tenth Street, Bloomington, IN 47405. E-mail: {sugimoto, guozhang, bcronin}@indiana.edu

Research on bias in peer review examines scholarly communication and funding processes to assess the epistemic and social legitimacy of the mechanisms by which knowledge communities vet and self-regulate their work. Despite vocal concerns, a closer look at the empirical and methodological limitations of research on bias raises questions about the existence and extent of many hypothesized forms of bias. In addition, the notion of bias is predicated on an implicit ideal that, once articulated, raises questions about the normative implications of research on bias in peer review. This review provides a brief description of the function, history, and scope of peer review; articulates and critiques the conception of bias unifying research on bias in peer review; characterizes and examines the empirical, methodological, and normative claims of bias in peer review research; and assesses possible alternatives to the status quo. We close by identifying ways to expand conceptions and studies of bias to contend with the complexity of social interactions among actors involved directly and indirectly in peer review.

Nature and Purpose of Peer Review

Peer review is an established component of professional practice, the academic reward system, and the scholarly publication process. The fundamental principle is straightforward: experts in a given domain appraise the professional performance, creativity, or quality of scientific work produced by others in their field or area of competence. In most cases, reviewer identity is hidden (single-blind review) to encourage frank commentary by protecting against possible reprisals by authors; and, in some cases, author identities will be masked from reviewers (double-blind review) to protect against forms of social bias. The structure of peer review is designed to encourage peer impartiality: typically, peer review involves the use of a "third party" (Smith, 2006, p. 178), someone who is neither affiliated directly with the reviewing entity (university, research council, academic journal, etc.) nor too closely associated with the person, unit, or institution being reviewed; and peers submit their reviews without, initially at least, knowledge of other reviewers' comments and recommendations. In some cases, however, peers will be known to one another, as with in vivo review, and may even be able to confer and compare their evaluations (e.g., members of a National Science Foundation [NSF] review panel).

Peer review, broadly construed, covers a wide spectrum of activities, including but not limited to observation of peers' clinical practice; assessment of colleagues' classroom teaching abilities; evaluation by experts of research grant and fellowship applications submitted to federal and other funding agencies; review by both editors and external referees of articles submitted to scholarly journals; rating of papers and posters submitted to conferences by program committee chairs and members; evaluation of book proposals submitted to university and commercial presses by in-house editors and external readers; and assessments of the quality, applicability, and interpretability of data sets (Lawrence, Jones, Matthews, Pepler, & Callaghan, 2011; Parsons, Duerr, & Minster, 2010). To this list one might add promotion and tenure decisions in higher education for which an individual's institutional peers and select outside experts determine that person's suitability for tenure and/or promotion in rank, and also the procedures whereby candidates are admitted to national academies, elected fellows of learned societies, or awarded honors such as the Fields Medal or Nobel Prize.

Received May 10, 2012; revised July 12, 2012; accepted July 13, 2012

 • Published online 6 December 2012 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.22784

Source (l): https://scholarlykitchen.sspnet.org/2011/12/08/is-peer-review-a-coin-toss/

Source (r): http://onlinelibrary.wiley.com/doi/10.1002/asi.22784/abstract

Scientific knowledge

Scientific knowledge is a portrait we paint of
(our understanding of) reality.

Necessary postulates for scientific working

- There are certain rules and principles for scientific working
- There is a scientific community to judge about the quality of scientific work
- There is a reality that exists independently of individuals' observations — the physical truth ("realism") — and individuals can make observations about (an excerpt of) reality
- Although observations may be faulty, it is possible (on the long run) to make reliable observations and to falsify incorrect statements about reality

Is Software Engineering research science?

What do you think?

Science can have different purposes

Basic Science ⟷ **Design Science**

Basic Science

- **Gaining** and **validating new insights**
- Often theoretical character
  Typically addressed by **natural and social sciences**

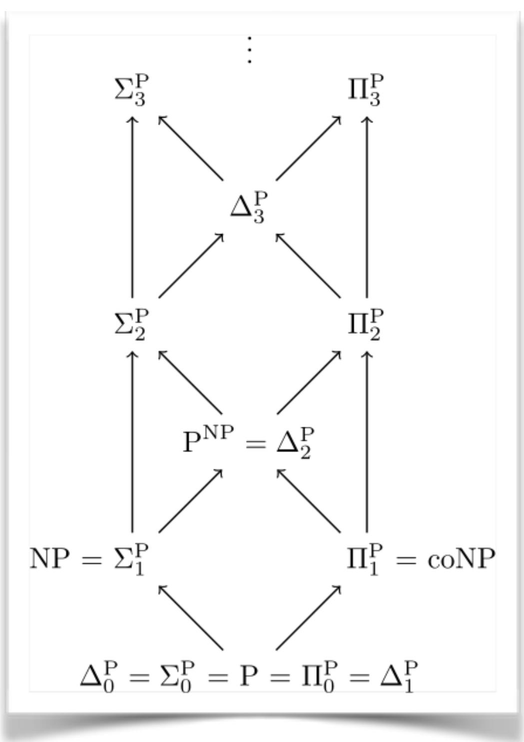
Design Science

- Guiding the application of scientific methods to **practical ends**
- Often rather practical (and pragmatic) character
  Typically addressed by **engineering disciplines**

In software engineering (research),
- we apply scientific methods to practical ends (treating design science problems)
- we also treat insight-oriented questions, thus, we are an insight-oriented science, too.

Science can have different purposes

Basic Science ⟷ Design Science



* Polynomial time hierarchy
(structural complexity theory)

Fundamental Research ⟷ Applied Research

Image Sources (left to right): Wikipedia, nasa.gov, Apple

Science can have different purposes

Basic Science ⟷ Design Science

Fundamental Research ⟷ *Applied Research*

Typically having more **“theoretical impact/relevance”**

Typically having more **“practical impact/relevance”**

Is Software Engineering research science?

Yes.

Outline

- Science (in a Nutshell)
- **Philosophy of Science - a Historical Perspective**
- Key Take Aways
- From Philosophy of Science to Empirical Software Engineering
- Empirical Software Engineering Processes
- Current Challenges in Empirical Software Engineering

What are Theories?

(A quick prologue)

Theories (generally speaking)

A **theory** is a belief that there is a pattern in phenomena.

Examples:

- Global warming was invented by the Chinese government to harm the US industry
- Vaccinations lead to autism

Are these theories scientific?

Speculations based on imagination or opinions that cannot be refuted

Further examples: https://twitter.com/realdonaldtrump

Scientific theories

A **scientific theory** is a belief that there is a pattern in phenomena while having survived
1. tests against experiences
2. criticism by critical peers

1. Tests

- Possibly experiment, simulation, trials
- Replication

2. Criticism

- Anonymous peer review / acceptance in the community
- Corroboration / extensions with further theories

Scientific theories

A **scientific theory** is a belief that there is a pattern in phenomena while having survived
1. tests against experiences
2. criticism by critical peers

Addresses so-called **Demarcation Problem** (distinguishing science from non-science)

1. Tests

- Possibly experiment, simulation, trials
- Replication

2. Criticism

- Anonymous peer review / acceptance in the community
- Corroboration / extensions with further theories

Scientific theories have…

A purpose

	Analytical	Explanatory	Predictive	Explanatory & Predictive
Scope	• Descriptions and con-ceptualisations, including taxonomies, classifications, and ontologies - What is?	• Identification of phenomena by identifying causes, mechanisms or reasons - Why is?	• Prediction of what will happen in the future - What will happen?	• Prediction of what will happen in the future and explanation - What will happen and why?

Quality criteria

- Testability
- Empirical support / (high) level of evidence
- Explanatory power
- Usefulness to researchers and / or practitioners
- …

Based on: Sjøberg, D., Dybå, T., Anda, B., Hannay, J. Building Theories in Software Engineering, 2010.

Scientific theories have…

By the way

Many theories in software engineering are so-called "design [science] theories", i.e. scientific theories about artefacts in a context.

[Artefact specification] x [Context assumptions] → [Effects]

More here: https://goo.gl/SQQwxt

A purpose

	Analytical	Explanatory	Pre	
Scope	• Descriptions and con-ceptualisations, including taxonomies, classifications, and ontologies - What is?	• Identification of phenomena by identifying causes, mechanisms or reasons - Why is?	• Prediction of what will happen in the future - What will happen?	• Prediction of what will happen in the future and explanation - What will happen and why?

Quality criteria

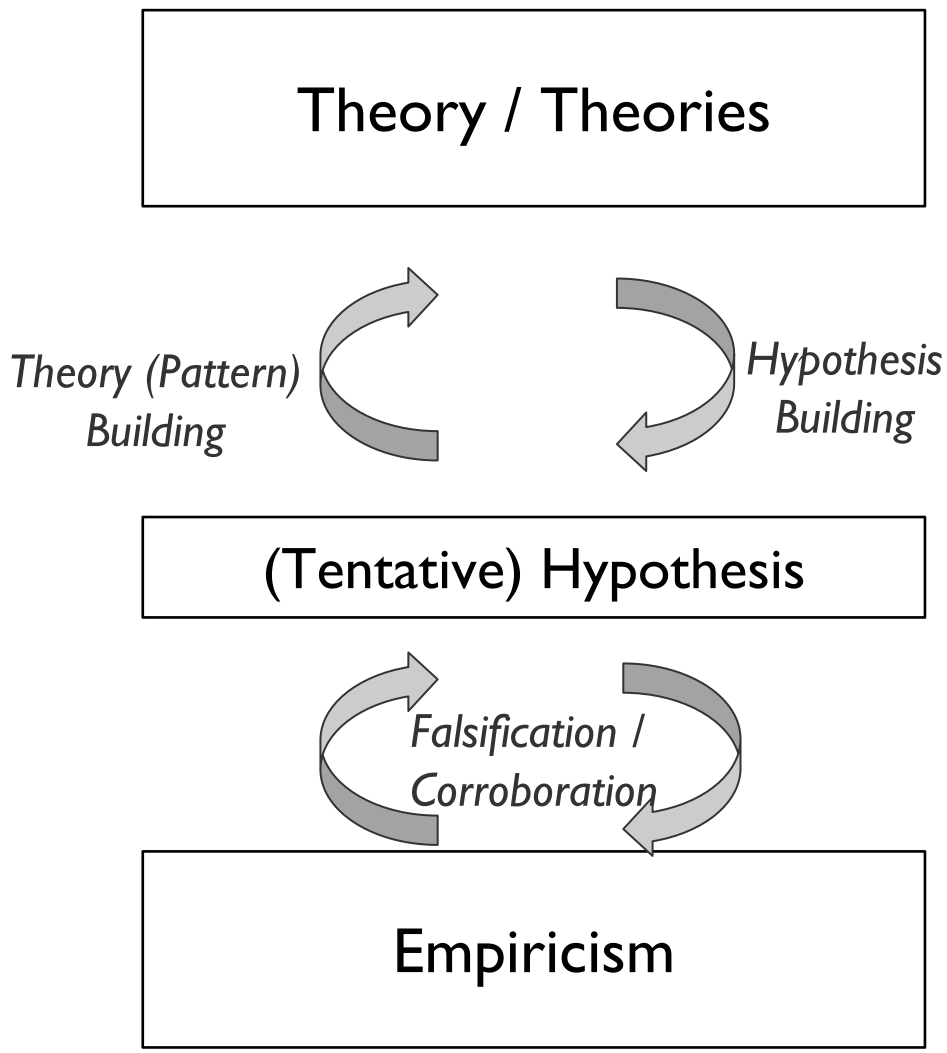
- Testability
- Empirical support / (high) level of evidence
- Explanatory power
- Usefulness to researchers and / or practitioners
- …

Laws "versus" theories

A law is a purely descriptive theory about phenomena (without explanations), i.e. an analytical theory.

Based on: Sjøberg, D., Dybå, T., Anda, B., Hannay, J. Building Theories in Software Engineering, 2010.

Theories and hypotheses



Scientific theory

- "[…] based on hypotheses tested and verified multiple times by detached researchers" (J. Bortz and N. Döring, 2003)

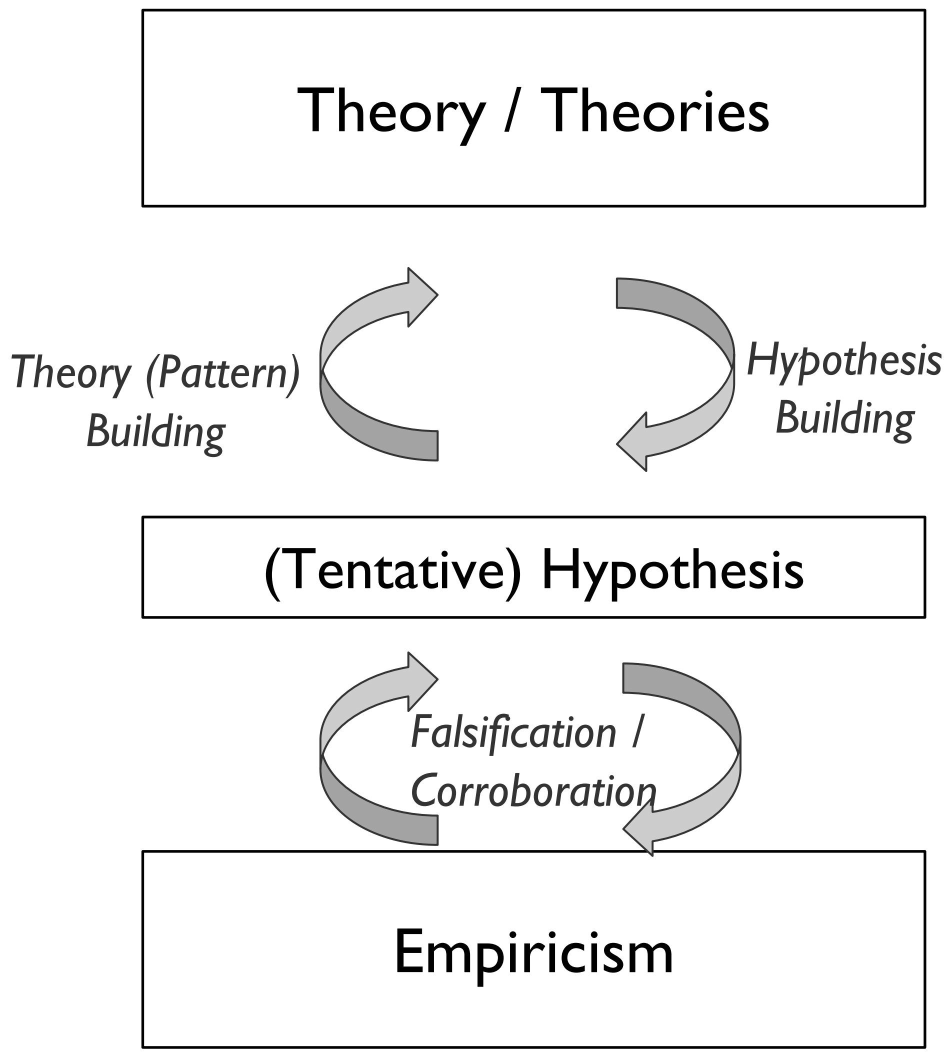
Hypothesis

- "[…] a statement that proposes a possible explanation to some phenomenon or event" (L. Given, 2008)
- Grounded in theory, testable and falsifiable
- Often quantified and written as a conditional statement

If cause/assumption (independent variables)
then (=>)
consequence (dependent variables)

Theories and hypotheses

By the way
We don't "test theories", but their consequences (via hypotheses)



Scientific theory

- "[…] based on hypotheses tested and verified multiple times by detached researchers" (J. Bortz and N. Döring, 2003)

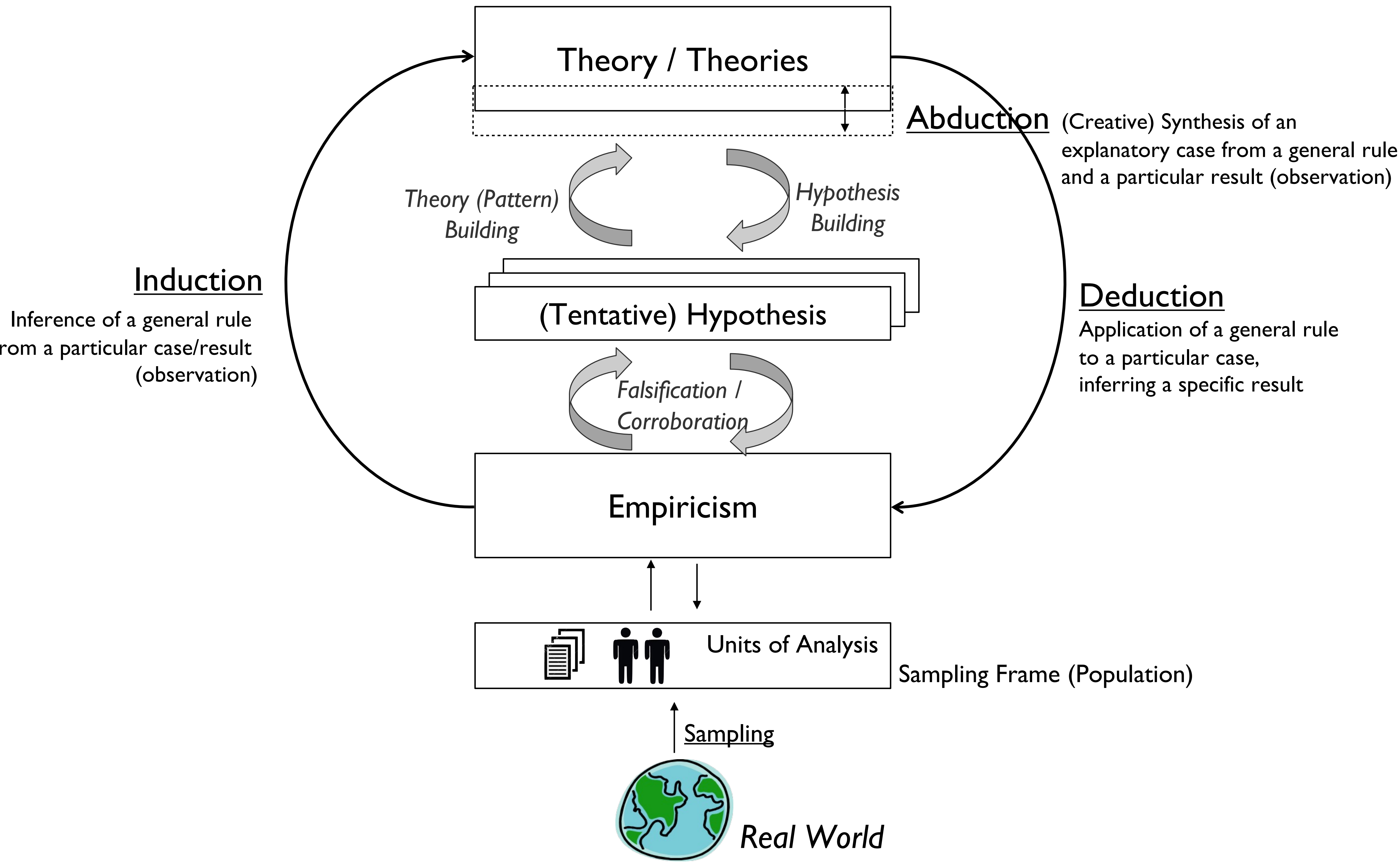
Hypothesis

- "[…] a statement that proposes a possible explanation to some phenomenon or event" (L. Given, 2008)
- Grounded in theory, testable and falsifiable
- Often quantified and written as a conditional statement

If cause/assumption (independent variables)
then (=>)
consequence (dependent variables)

From real world to theories… and back

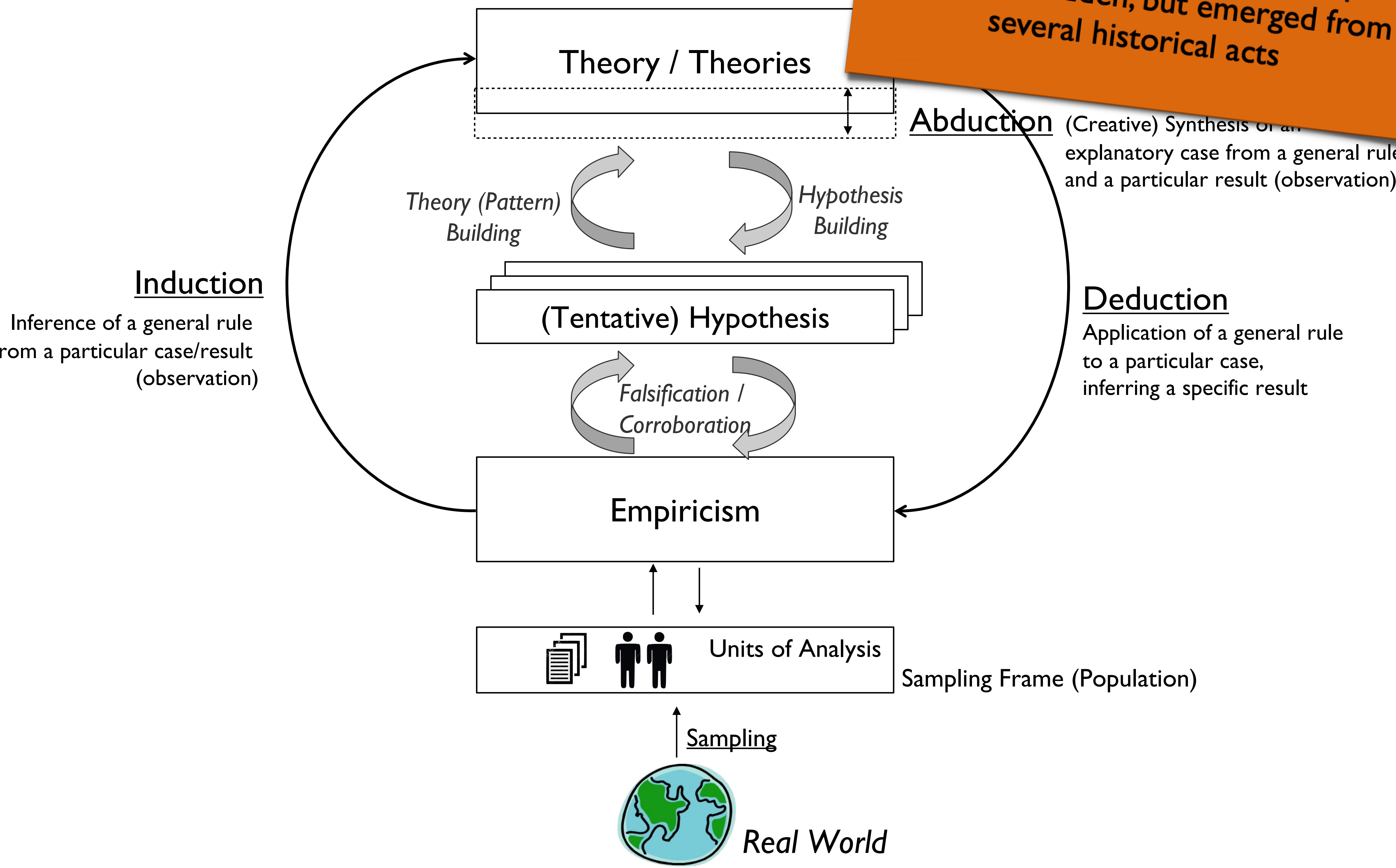
Principles, concepts, terms



From real world to theories… a

Principles, concepts, terms

This understanding wasn't developed out of the sudden, but emerged from several historical acts



An Introduction into the (History of) Philosophy of Science…

… in several Acts

Act 1

Era of Positivism

Gaining knowledge through sensory experiences

Image Source: Antoine de Saint-Exupéry. Le Petit Prince, 1943.

Origin and principles

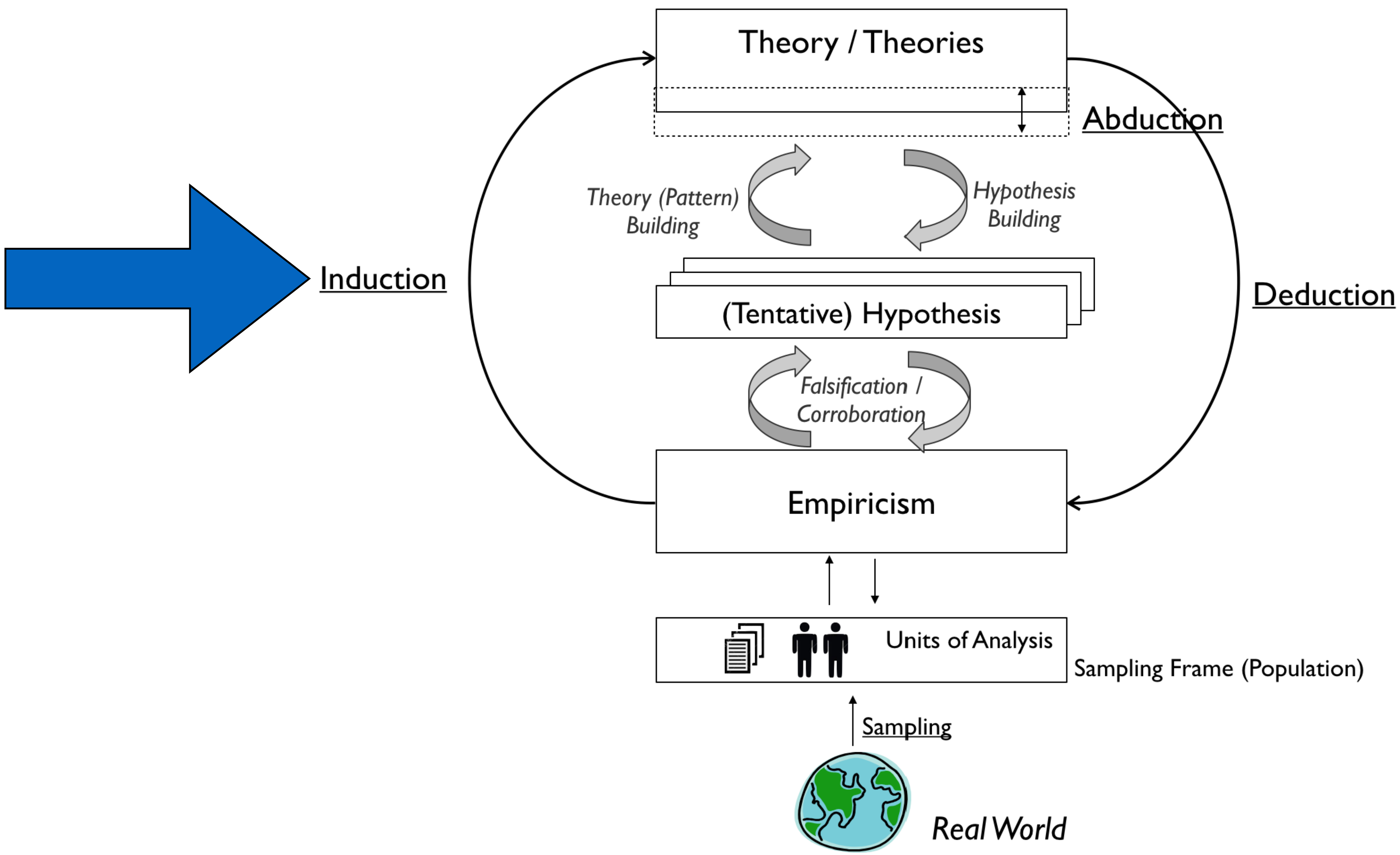
- Positivism traced back to Auguste Comte (1798–1857). (*A. A General View of Positivism, 1848 (French), 1865 (English)*.)
- Emerges from a secular-scientific ideology in response to European secularisation (Enlightenment - *Voltaire*)

Knowledge (i.e. theories)

- Must not be governed by its association with divine presences
- Derived from sensory experiences (based on empirical evidence)
- Interpreted through reason and logic
- Only source of truth

Scope

Knowledge growth through sensory experiences.



Example

Theory: “All Swans are white”

This statements (to be true) requires:

- Knowledge about whole universe of swans
  (which exist, which have existed, and which will exist)
- Objective interpretation of real world references

Limitations

1. Insufficient knowledge about the universe

Inductive inference consists of *generalisation* from observations made in some *finite sample* to *broader population of instances* (enumerative induction)

Finite set of observations is logically compatible with multitude of generalisations

2. Subjectivity in sensory experiences

Theories built upon underlying cognitive schemas and existing mental models

No amount of observations can (sufficiently) justify a universal belief

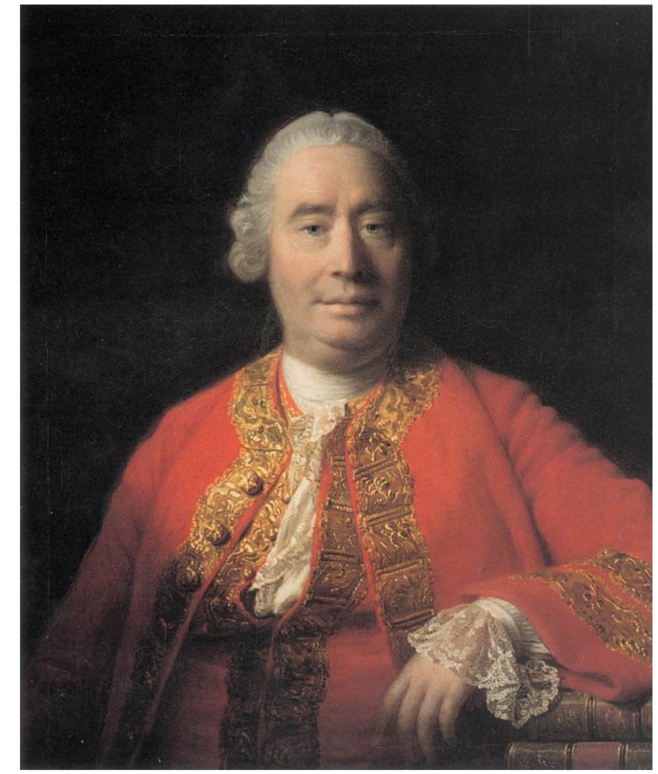
> *Induction is the glory of science and the scandal of philosophy.*
>
> — Broad, 1968

The problem with inductive reasoning is not per-se a problem of science (or scientific methods) so much as it is a problem of knowledge

Act 2

Era of Scientific Realism

Principle problem of "induced" knowledge



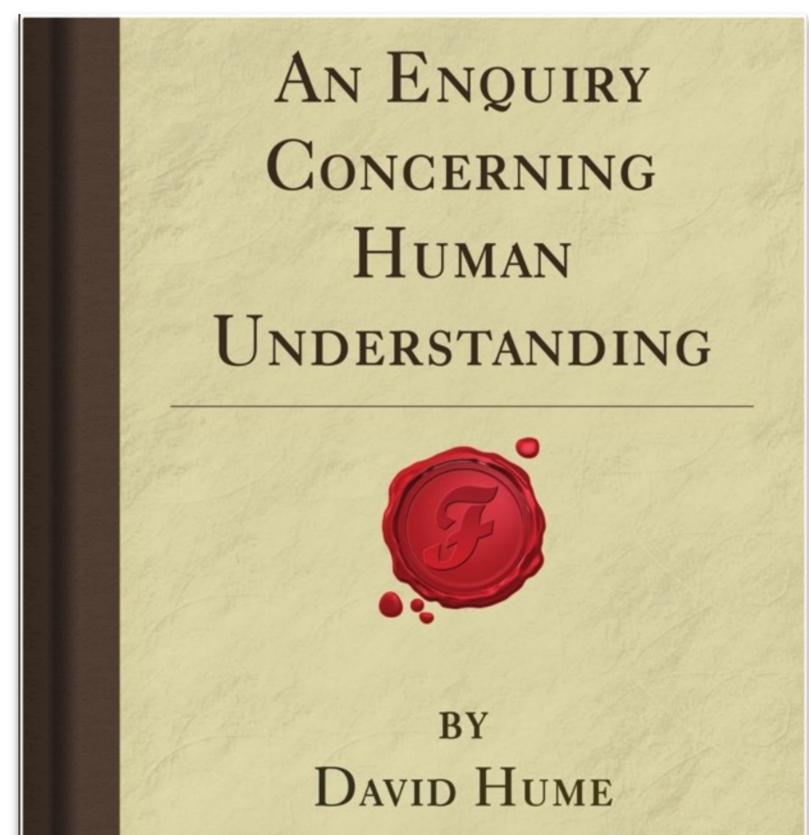
- David Hume (1711 — 1776) questions extent to which inductive reasoning can lead to knowledge

  Inductive reasoning alone (and belief in causality), cannot be justified rationally

Relation to (predictive) theory building

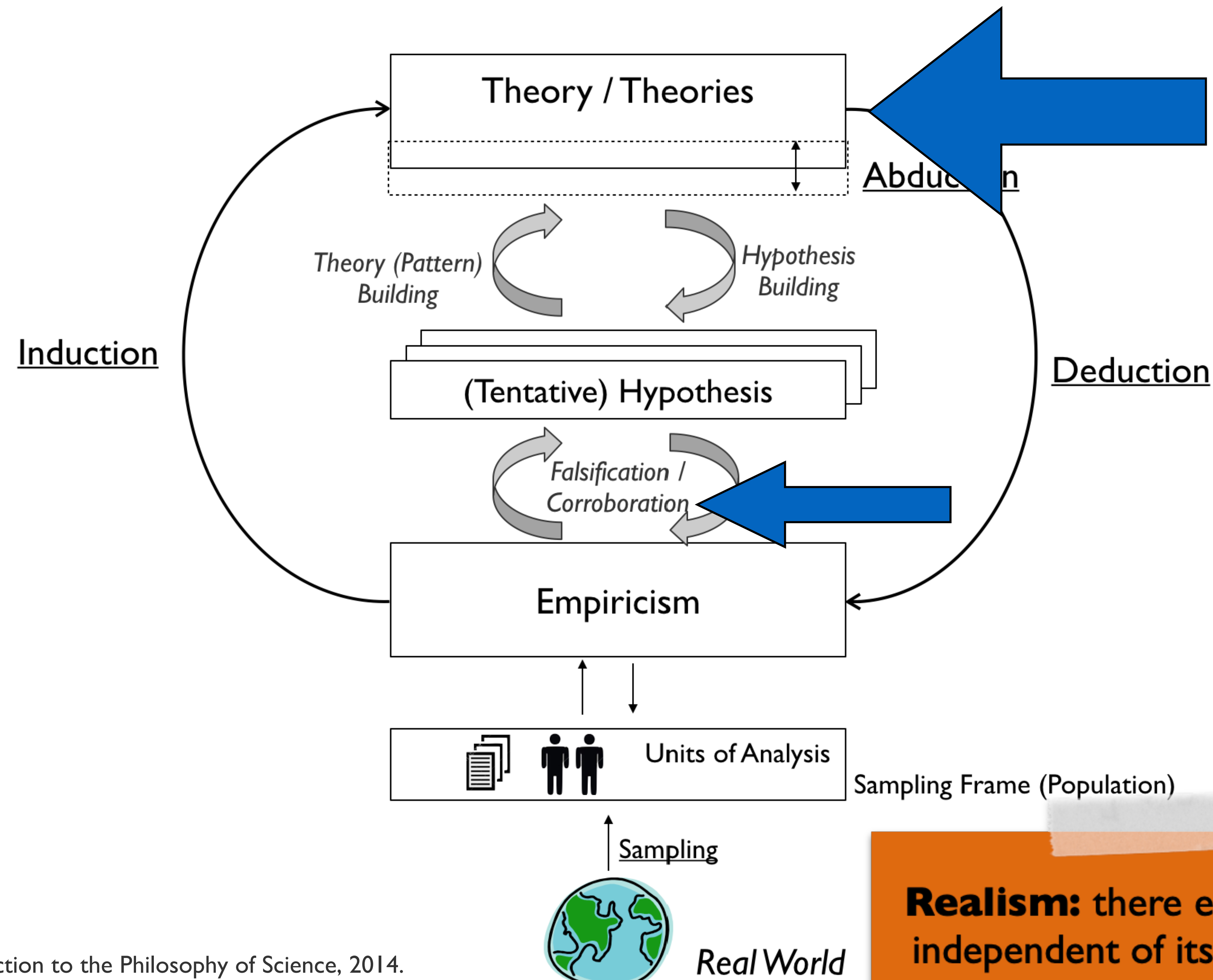
- Beliefs about future based on
  - experiences about the past and
  - assumption that the future will resemble the past
- **However,** thousands of observations of event A coinciding with event B do not allow to logically infer that all A events coincide with B events
- Example: It is logically possible that the sun won't rise tomorrow

  We don't *know* that the sun will rise tomorrow, yet it is *reasonable to believe* (to a certain extent) it will rise



Scope

Scientific theories are (probably) approximately true when they achieve a certain level of success in prediction and experimental testing.



Realism: there exists a reality independent of its observation

Based on: Staley,, K. An Introduction to the Philosophy of Science, 2014.

Excursion

Related: Bayesianism

- Traced back to Rev. Thomas Bayes 1701 – 1761 (essays published posthumously by Richard Price, then popularised by Pierre-Simon Laplace as today's *Bayesian probability*)
- Basis for theory of *rational belief* (on mathematical framework of probability theory)

Doctrine of chances (briefly)

- Method of calculating the probability of all conclusions founded [so far] via induction

Probabilities represent current state of belief ("knowledge") in light of currently available evidence

We "know" with certain confidence, i.e. strength of belief

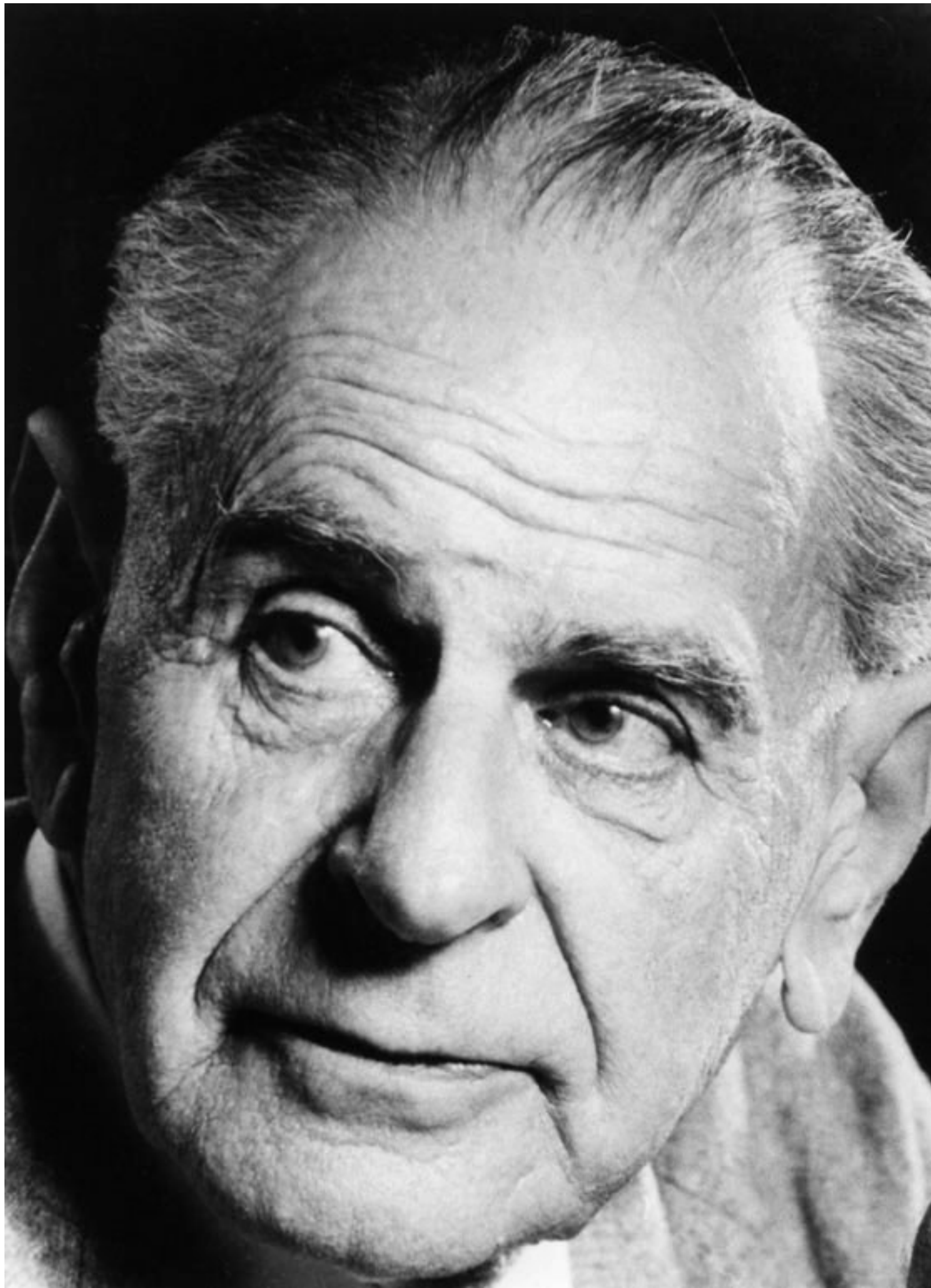
Act 3

Era of Critical Rationalism

Origin and principles

- Traced back to Sir Karl Popper (1902 - 1994).
- Popper sees problems in induction as so sever that he rejects it completely
- Response to logical positivism, i.e. verification by experience, as (initially) propagated by *Vienna Circle* (scientists meeting annually at the University of Vienna… and also at Café

> *"Positivism is as dead as a philosophical movement can be"*
>
> — Passmore

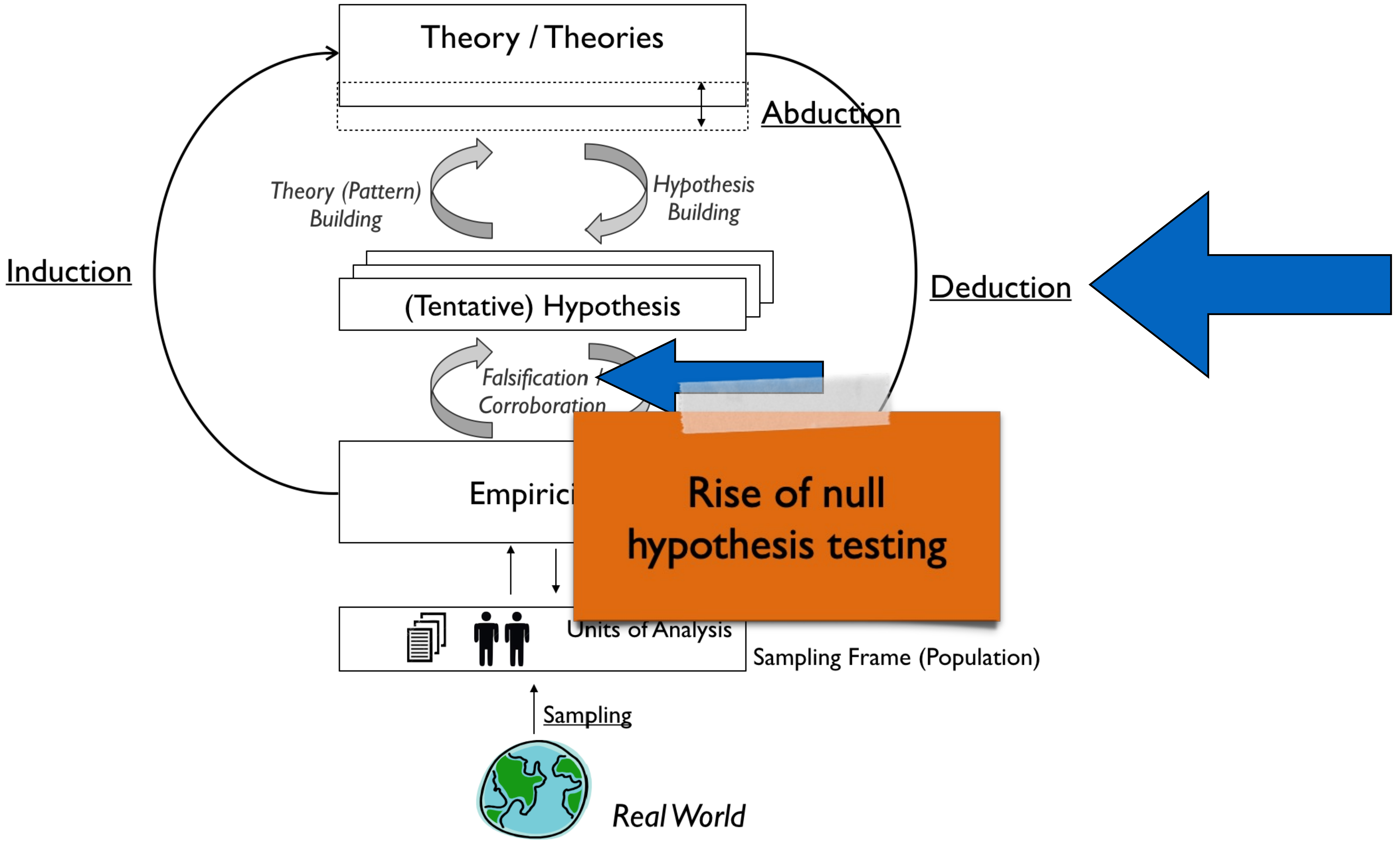


Falsification as demarcation criterion

- From *supporting theory via corroboration* to *criticising and refuting / rejecting* it
- Only falsifiable theories are scientific

Scope

Knowledge growth through falsification.



Principles for building and accepting theories

- **Falsifiability** centres not on what a hypothesis says will happen, but on what it forbids, i.e. on experimental results that should not be produced

  Always prefer those theories that are the most falsifiable ones (to have survived testing so far)

> *A more falsifiable theory "says more about the world of experience" that one that is less falsifiable because it rules out more possible experimental outcomes.*
>
> — Popper, 1992

- **Theories are** never solid, but they can be **sufficiently robust** to be commonly accepted **after standing strong and repetitive attempts for falsification**

  Robustness of theories not by support / corroboration (free of inductive valences), but by extent to which it has survived falsifications

Limitations of critical rationalism

If a theory cannot be refuted,
it may be also because:

1. One or more hypotheses are inadequate (if so, which one?)
2. "Underdetermination" problem
   - insufficient data
   - insufficient knowledge about causal relationships
3. Particularities of the context and conditions
4. Observations are incorrect
   - wrong or even not yet existing measurement
   - "wrong" interpretation

Often impossible to tell apart.

> *"[…] the physicist can never subject an isolated hypothesis to experimental test, but only a whole group of hypotheses* [and if the tests fail], *the experiment does not designate which one should be changed"*
>
> — Duhem, 1962

Rejection of statements depends on many non-trivial factors

Act 4

Era of (pragmatic) Constructivism

Pragmatism and Constructivism

Pragmatism is the recognition that there are many different ways of interpreting the world and undertaking research, that no single point of view can ever give the entire picture.

Rise of mixed research methods

Constructivism is the recognition that reality is a product of human intelligence interacting with experience in the real world.

> *As soon as you include human mental activity in the process of knowing reality, you have accepted constructivism.*
>
> — *Elkind, 2005*

Rise of qualitative research methods

Origin and principles

- P*ragmatism* initially coined by logician Charles Sanders Peirce (1839 – 1914)
- *Constructivism* initially coined by psychologist Jean Piaget (1896 – 1980)

Maxims

- **Pragmatism:** Method appropriateness judged by extent to which it answers inquiry question at hand

  Value of methods (and theories) depends also on practical usefulness to solve a problem (W. James)

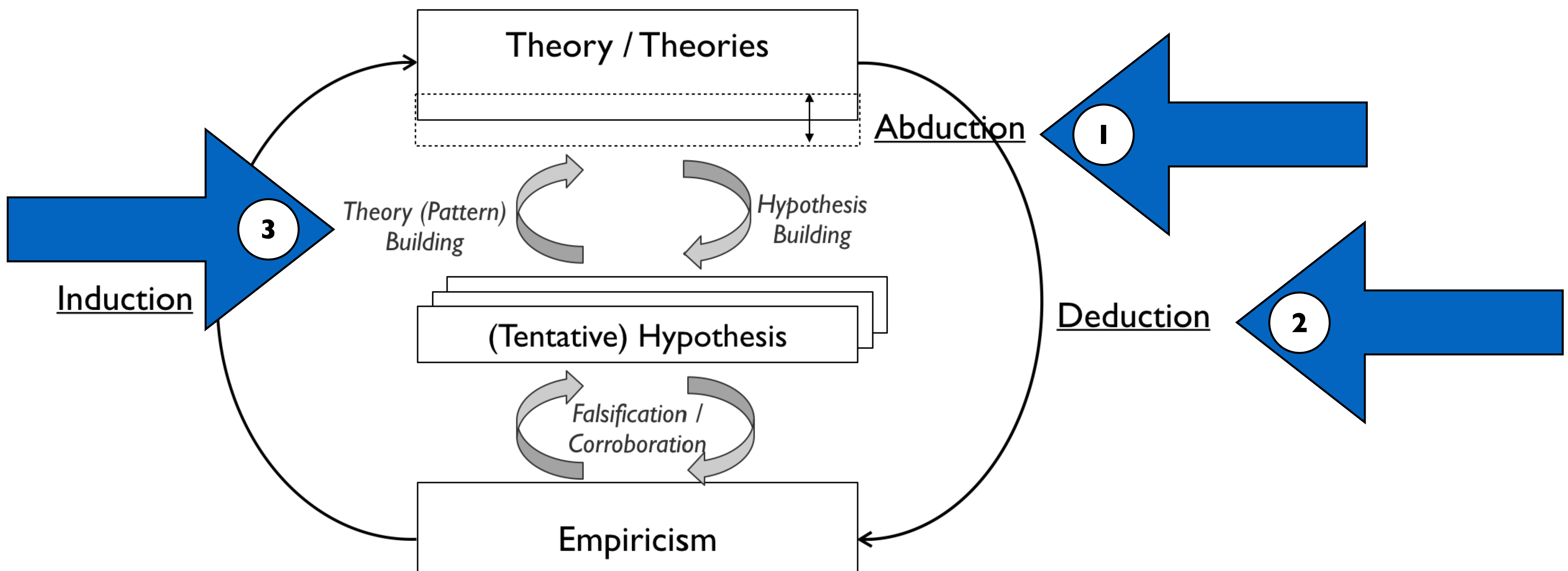
- **Constructivism:** Accept that theories, background, knowledge and values of the researcher influence interpretation of physical reality

  Scientific working is also a creative task

  "Truth" depends (also) on acceptance by those who interpret reality

Scope

Knowledge growth comes in an iterative, step-wise manner* where researchers also may (or must) leave the realms of logic and apply creative reasoning.



*** Approach as introduced by Peirce**

1. Identify hypothesis via abduction
2. Deduce consequences
3. Induce further facts to support hypothesis (otherwise return to 1.)

From Rationalism to Pragmatism

	Rationalism Quantitative research	Constructivism Qualitative research	Pragmatism Mixed method research
What is the relationship between researcher and subject/object?	Researcher independent from what is being researched	Subjects interpret their "own" reality, researcher can become insider	
What is the research strategy?	**Deductive** • Hypothesis testing (corroboration / falsification) • Context free • Generalisations for predicting, explaining, and understanding	**Inductive** • (Active) theory building • Context bound • Patterns and theories for understanding	**Combination of inductive and deductive** • Context bound • Patterns and theories for understanding • Generalisations for predicting and explaining

What happened so far?

Local Problem-Solving View

1. **Positivists (and realists)** infer scientific knowledge **-** at least with a certain level of confidence - from direct observations (but what is this?)
2. **Rationalists** replace worse by better theories using falsification (but it is often unclear where problems lie; in the theory or in the observation?)
3. **(Pragmatic) constructivist** add a creative (and pragmatic) perspective for an iterative and local problem-solving

How does science progress in the long run?

Act 5

Era of Post-Positivism

The empirical basis of objective science has nothing 'absolute' about it. *Science does not rest upon solid bedrock.The bold structure of its theories rise, as it were, above the swamp. It is like a building erected on piles. The piles are driven down from above into the swamp, but not down to any natural or 'given' base; and* ***if we stop driving*** *the piles* ***deeper, it is not because we have reached firm ground. We simply stop when we are satisfied*** *that the piles are firm enough to carry the structure,* ***at least for the time being.***

— Popper, 1992

Origin and principles

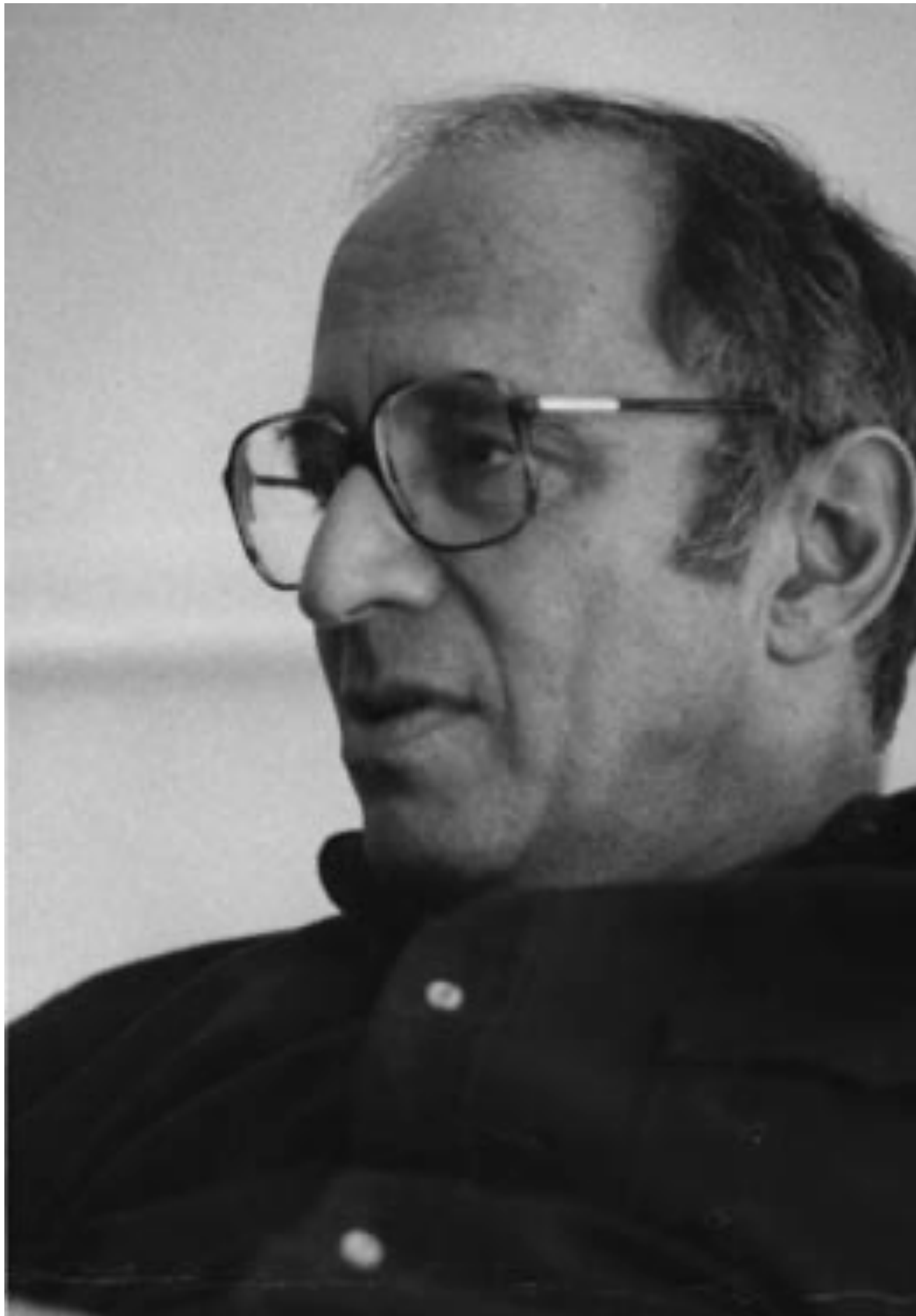
- Initially coined by Thomas Kuhn (1922-1996)
- Scientific progress doesn't follow piecemeal falsification / corroboration, but is revolutionary and influenced by sociological characteristics of scientific communities.
- Scientists work within paradigms (and are uncritical towards their paradigm)

Maxim of paradigms

- *Paradigm* is set of accepted fundamental laws, assumptions, standard ways of working (instrumentation and techniques)

  Normal scientific activity is a puzzle-solving activity. Failures are failures of scientists, not the paradigm; puzzles that resist solution are usually anomalies rather than falsifications.

  Progress via "revolutionary paradigm shift"



Scientific progress via “paradigm shifts”

1. Scientists work in communities within certain (incommensurable) paradigms
2. If no progress can be observed, it is an indicator for a crisis
3. A change of paradigm (“paradigm shift”) by acceptance of the community

Acceptance first, arguments later

Examples

- Copernican revolution
- Development of quantum mechanics
- Agile methods?

Limitation

No notion of when a paradigm is „better“ than another

“[…] judging a theory by assessing the number, faith, and vocal energy of its supporters […] basic political credo of contemporary religious maniacs”
— Lakatos, 1970

At the moment physics is again terribly confused. In any case, it’s too difficult for me, and I wish I had been a movie comedian or something of the sort and had never heard of physics.
— Kronig, 1960

Though the world does not change with a change of paradigms, the scientist afterwards works in a different world.
— Kuhn. The Structure of Scientific Revolutions, 1962

Research programmes

- Coined by Imre Lakatos (born as "Lipschitz")(1922-1974)
- Kuhn's revolutionary science had no notion of when a paradigm is „better" than another, i.e. often not clear which hypothesis in a structure of hypotheses (i.e. theory) problematic

Structure via research programmes

- *Hard core:* Non-falsifiable
- *Protective belt:* falsifiable

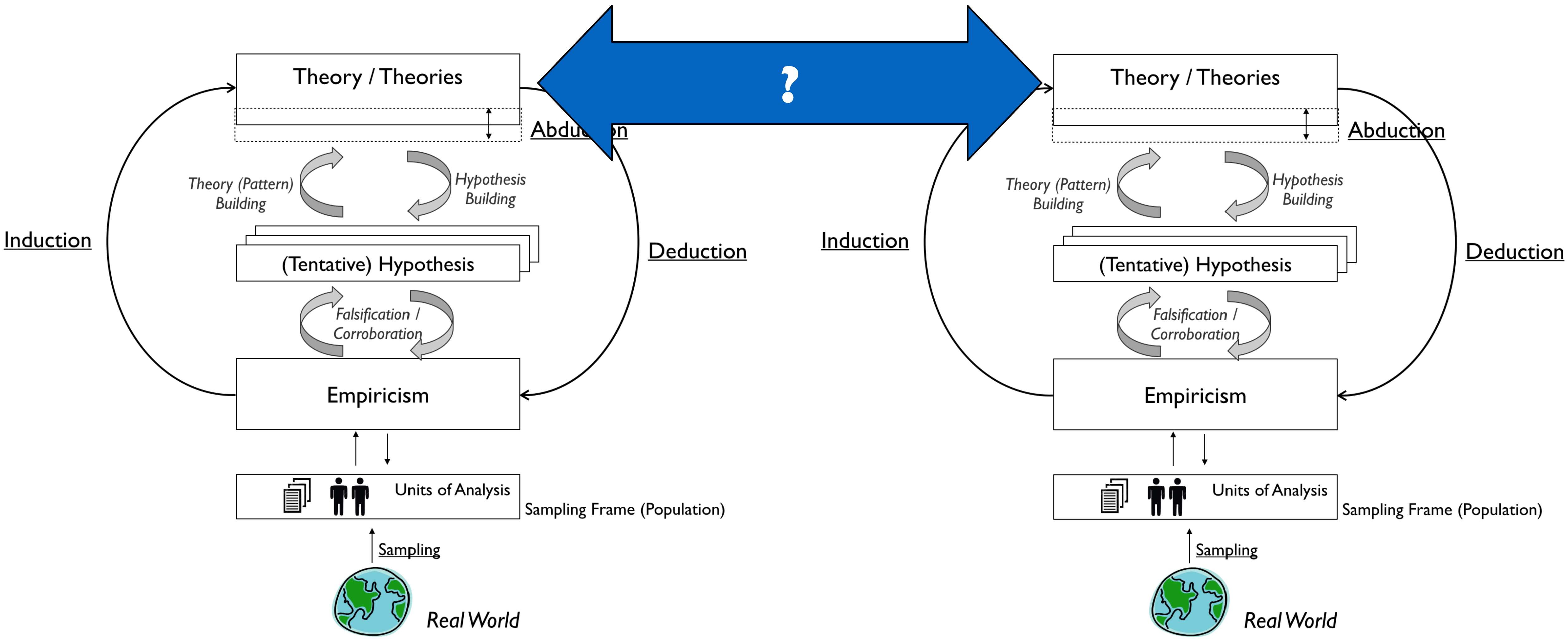
Progress by modifying protective belt in testable way: Progressive research over degenerating research

Degenerative research: explaining what is already known

Progressive research: based on ability to predict novel facts

Scope

Knowledge growth not by following the (piece-wise) falsificationist or inductionist approaches, but through (in parts competing) programmes.



Limitations

1. No applicability to local problem-solving

- Paradigm / programme debates not about (relative) problem-solving ability, but about which paradigm should in future guide research on problems (such a decision made based on faith)

  No support for “quick wins” as (e.g.) in falsification as novelty can only be seen after a long period of (competing) programmes and continuous work within those programmes

  (Still helps understanding social mechanisms involved)

2. Advancing knowledge is a paradigm/programme debate

- Relies on acceptances by the communities based of belief to which extent theories can solve existing and future problems (science comes along a social and sometimes political process)

  Progress based on acceptance by protagonists in communities

Act 6

Era of Epistemological Anarchy

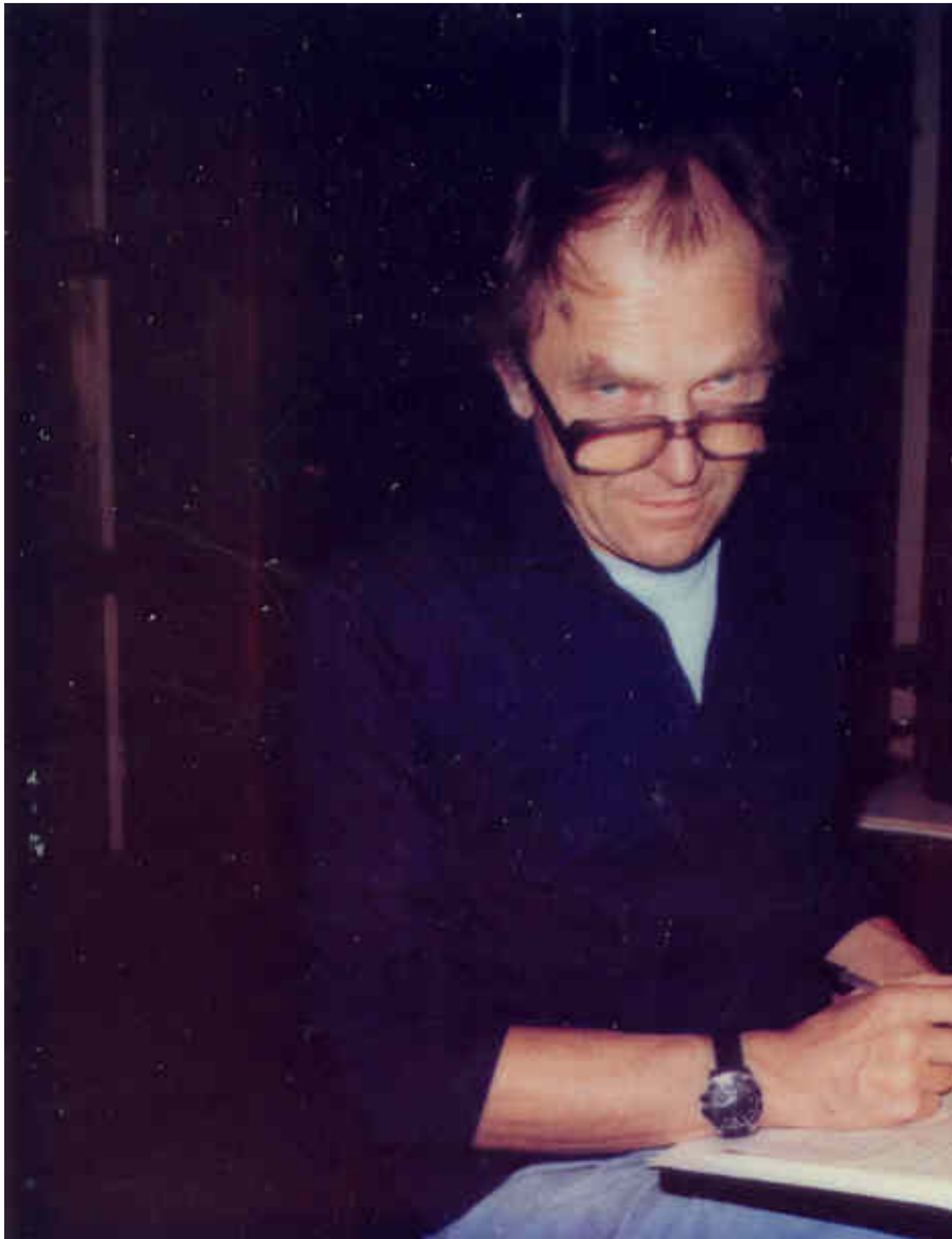
Origin and principles

- Coined by Paul K. Feyerabend (1924-1994)
- Did not express own conviction, but provoked communities to question theirs

Maxim of “Anything Goes”

- Reject idea that there can be a universal notion of science (at least without ending up in total relativism)
- Reject any attempt to constrain science by acceptance as it
  - inhibits free development of individual scientist
  - blocks growth of scientific knowledge

  Chose whatever others might think is „progress“ and play the devil’s advocate



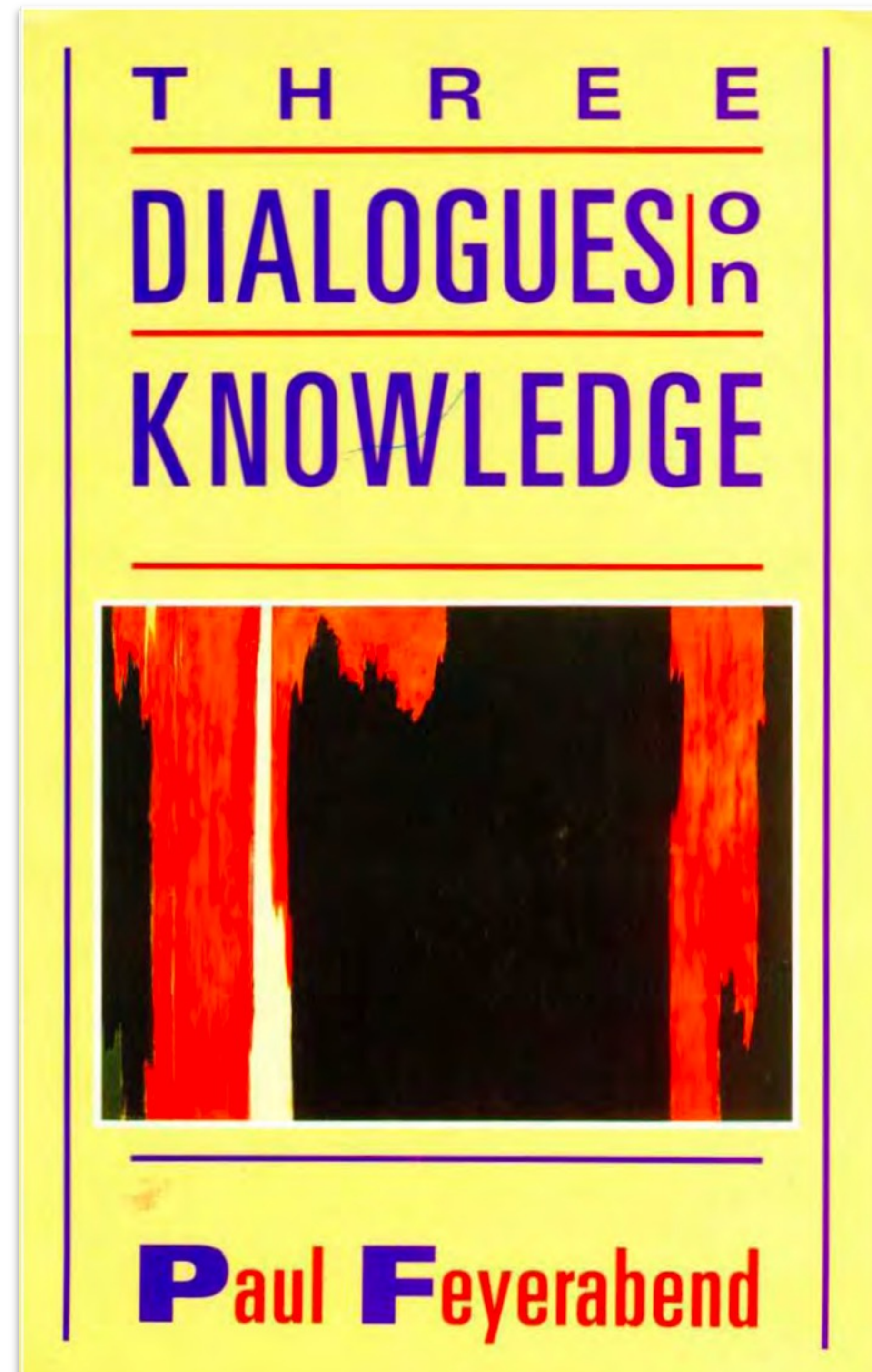
Paul Feyerabend: The (polemic) Devil's Advocate

Paul Feyerabend, also known as the

- Defender of Creationism
- Defender of Astrology

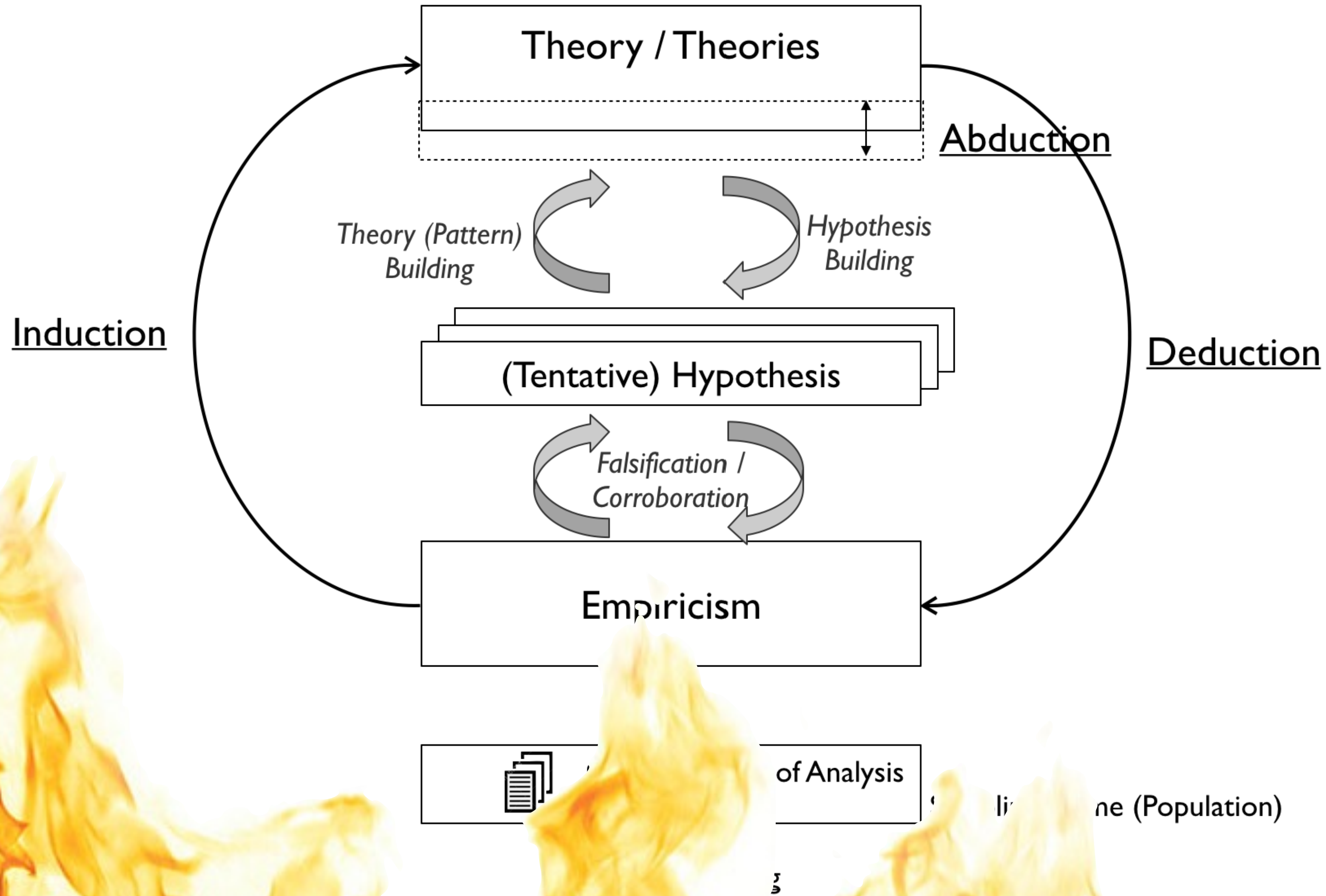
> *Astrology bores me to tears [, but] it was attacked by scientists, Nobel Prize winners among them, without arguments, simply by a show of authority and in this respect deserved a defence.*
>
> — Feyerabend, 1991

Devil's advocate



Scope

Knowledge growth by introducing new theories that challenge the established facts of any given time (“anything goes”).



Principle: Reject authorities and challenge what we accept as "factually known"

1. No such thing as universal way of scientific working

- Any rule used as "universal guide" to scientific working might, under some circumstances, prevent scientists from contributing to the progress of science

  "Keep our options open"

> *Effectiveness of a rule for pursuing science depends on what the world is like which is exactly what we do not know.*
>
> — Feyerabend (via K. Staley)

2. No such thing as (universally acceptable) truth

- Every explanation (no matter how absurd) is possible for an observation
- No authority should be accepted

  The highest duty of a scientist is to play the devil's advocate

In which era do we live today?

Ideally, in all of them.

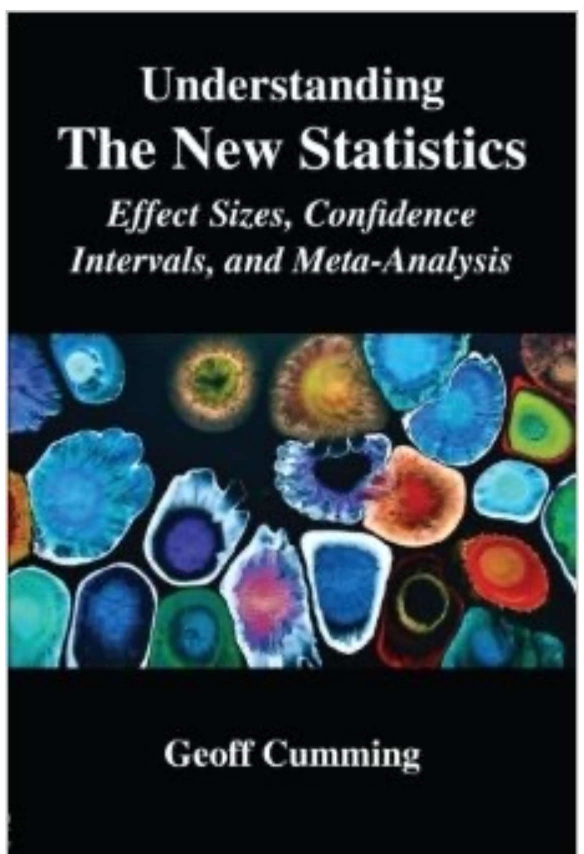
All views and contributions need to be considered

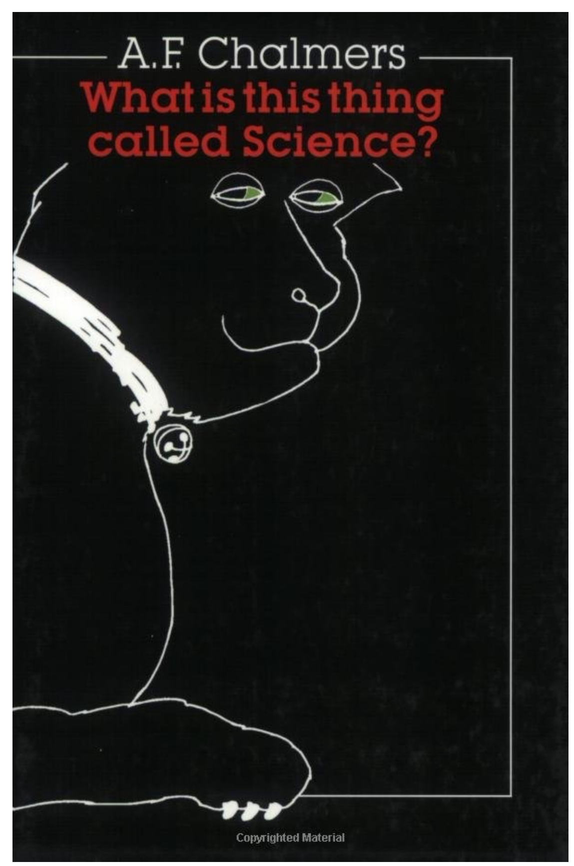
There is not the one "correct" epistemological approach, but many lessons we can learn from their historical evolution.

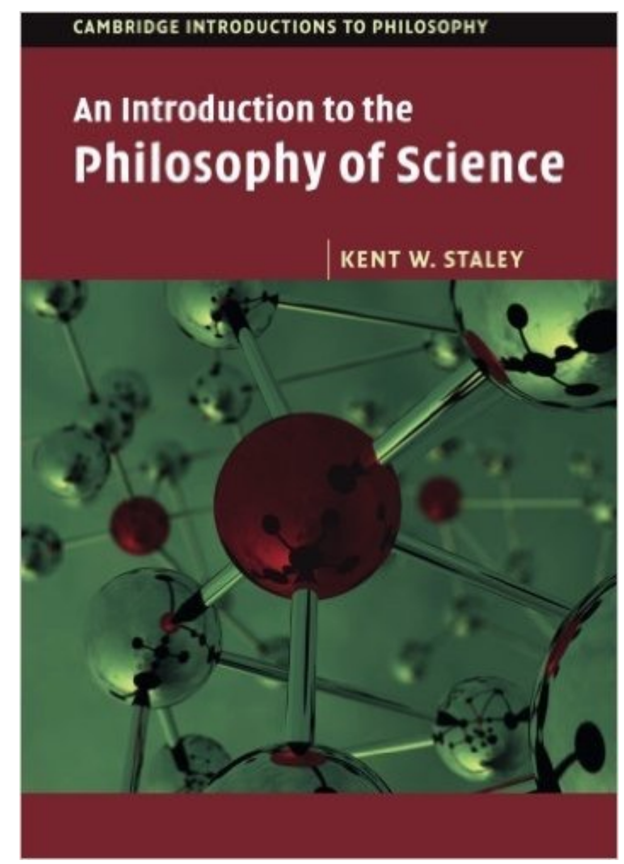
Further reading

Introduction into (one) current debate

Overview of movements and their historical context







(Many quotes based on this book)

Outline

- Science (in a Nutshell)
- Philosophy of Science - a Historical Perspective
- **Key Take Aways**
- From Philosophy of Science to Empirical Software Engineering
- Empirical Software Engineering Processes
- Current Challenges in Empirical Software Engineering

What are your take-away(s)?

Beware the basic principles of scientific progress

1. No such thing as absolute and / or universal truth (truth is always relative)

2. The value of scientific theories always depends on their
   - falsifiability,
   - ability to stand criticism by the (research) community,
   - robustness / our confidence (e.g. degree of corroboration),
   - contribution to the body of knowledge (relation to existing evidence), and
   - ability to solve a problem (e.g. practical problem).

3. Theory building is a long endeavour where
   - progress comes in an iterative, step-wise manner,
   - empirical inquiries need to consider many non-trivial factors,
   - we often need to rely on pragmatism and creativity, and where
   - we depend on acceptance by peers (research communities)

4. Scepticism and also openness are major drivers for scientific progress

Image Source: Monty Python

Adopt fundamental credos of scientific working

1. Be sceptical and open at the same time:
    - no statement imposed by authorities shall be immune to criticism
    - be open to existing evidence and arguments/explanations by others

2. Be always aware of
    - strengths & limitations of single research methods
    - strength of belief in observations (and conclusions drawn)
    - validity and scope of observations and related theories
    - relation to existing body of knowledge / existing evidence

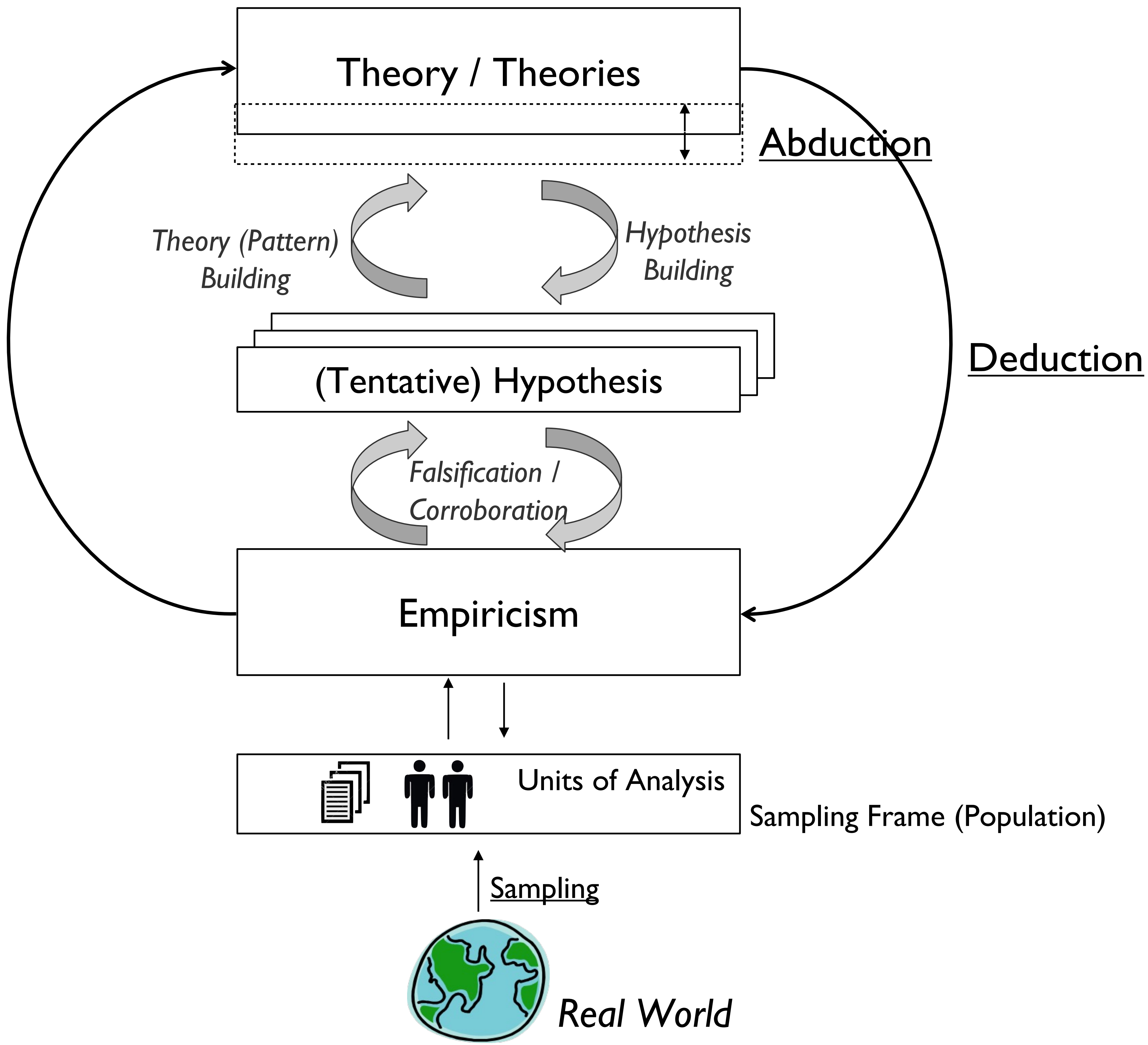
3. Appreciate the value of
    - all research processes and methods
    - null results (one’s failure can be another one’s success)
    - replication studies (progress comes via repetitive steps)

4. Be an active part of something bigger (knowledge is built by communities)

Image Source: Monty Python

Understand the research methods: their purposes, strengths, limitations, and places in a bigger picture

- Ethnographic studies, Folklore gathering
- Case studies / Field studies (Exploratory)
- Case studies / Field studies (Confirmatory)
- Formal analysis / logical reasoning
- Grounded theory



And yet, too often we s

An **understanding** of the **foundations and implications of scientific methods** is crucial for building a reliable body of knowledge (via **theories**) in our field.

Research Question: Which car has the best driving performance?
H_0: There is no difference.

20 people without a driving licence participated.
We taught them to drive in a lecture of 2 hours.

Results: The BMW is significantly better than the Daimler. ($p<0.01$)

Adapted from: Dag I.K. Sjøberg (University of Oslo) Keynote at the International Conference on Product-Focused SW Process Improvement 2016, Trondheim, Norway.
Image Sources: Company websites

Outline

- Science (in a Nutshell)
- Philosophy of Science - a Historical Perspective
- Key Take Aways
- **From Philosophy of Science to Empirical Software Engineering**
- Empirical Software Engineering Processes
- Current Challenges in Empirical Software Engineering

Software Engineering research

- Software engineering is development (not production), inherently complex, and human-centric

  (Empirical) research methods allow us to
  - Reason about the discipline and (e.g. social) phenomena involved
  - Recognise and understand limits and effects of artefacts (technologies, techniques, processes, models, etc.) in their contexts

Exemplary questions

- There exist over 200 documented requirements engineering approaches
  - Which one(s) work in my context?
  - To which extent? Under which conditions?
- There is a new method for requirements elicitation
  - What are the strengths and limitations?

Building a reliable body of knowledge (theory building and evaluation) is key for progress in our field.

Empirical Software Engineering

The **ultimate goal of empirical Software Engineering** processes is **theory building and evaluation** to strengthen and advance our body of knowledge.

Practitioners “versus” Researchers

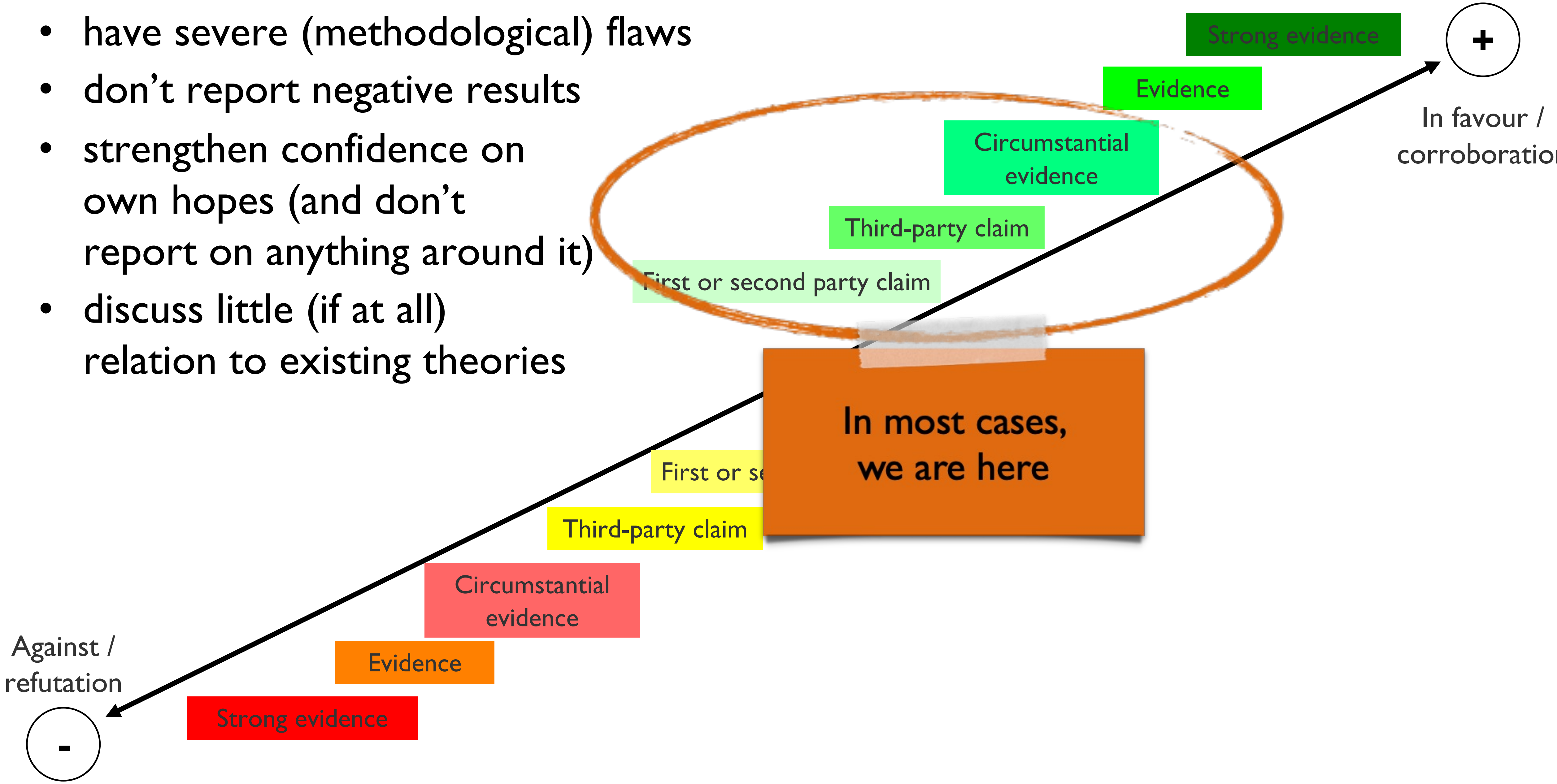
- Researchers usually concerned with understanding the nature of artefacts and their relationship in the context
  - *What is the effect?*
  - *Why is it so?*
- Practitioners usually concerned with improving their engineering tasks and outcomes, using available knowledge
  - *What is the problem?*
  - *What is the best solution?*

In our field, **theoretical and practical relevance** have often a special **symbiotic relation**

Current state of evidence in Software Engineering

Available studies too often

- have severe (methodological) flaws
- don't report negative results
- strengthen confidence on own hopes (and don't report on anything around it)
- discuss little (if at all) relation to existing theories



Source [for levels of evidence]: Wohlin. An Evidence Profile for Software Engineering Research and Practice, 2013.

Current state of evidence in Software Engineering

The current state of empirical evidence in Software Engineering
is still weak.

- We still lack robust scientific theories (let alone holistic ones)
- Symptom: Many movements based on conventional wisdom, e.g.:
  - #noestimates (look it up on Twitter ;-)
  - goal-oriented requirements engineering (to be taken with a grain of salt)

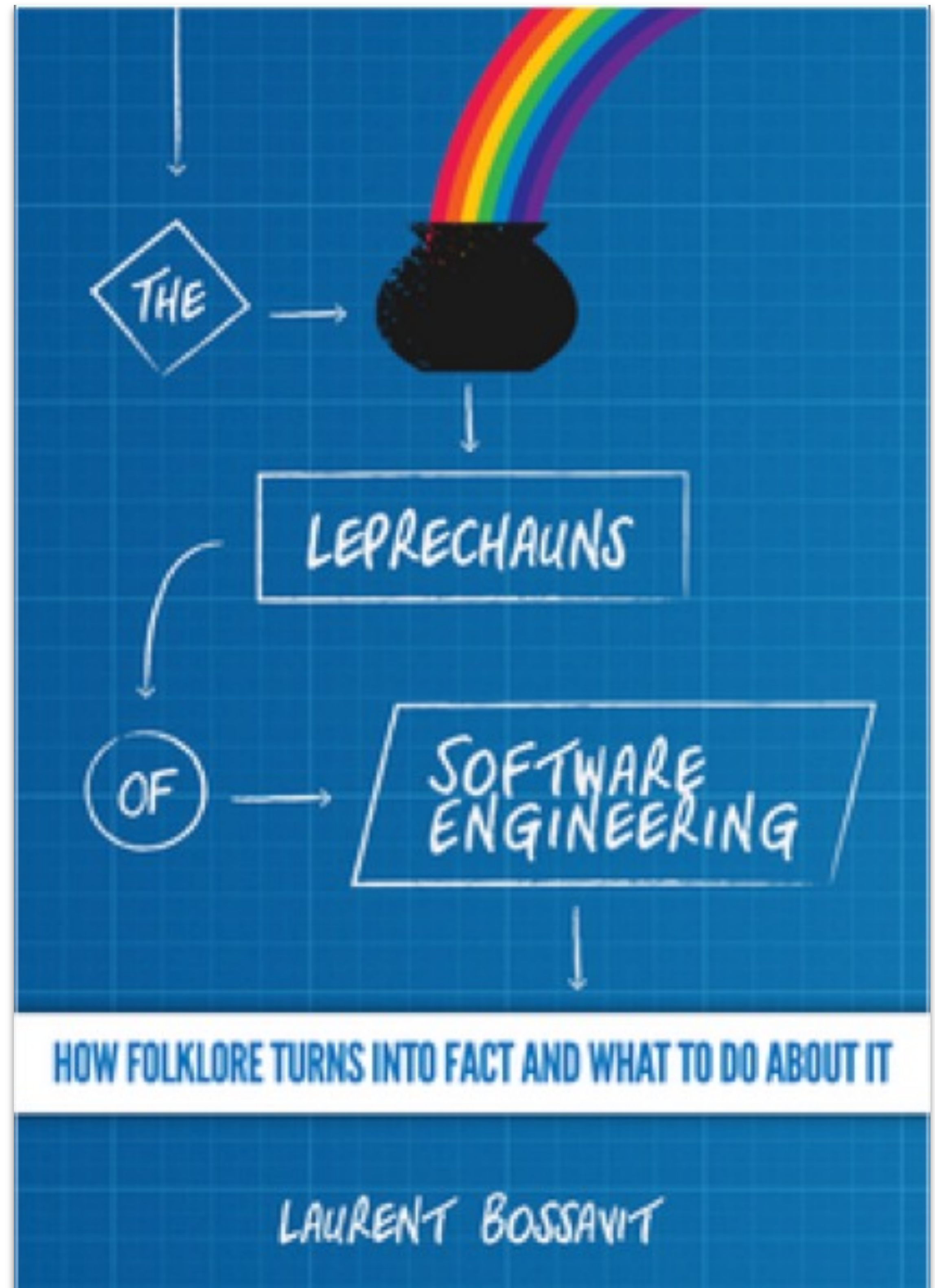
  Software engineering is, in fact, dominated by many "Leprechauns"

Leprechauns of Software Engineering

Folklore turned into “facts”

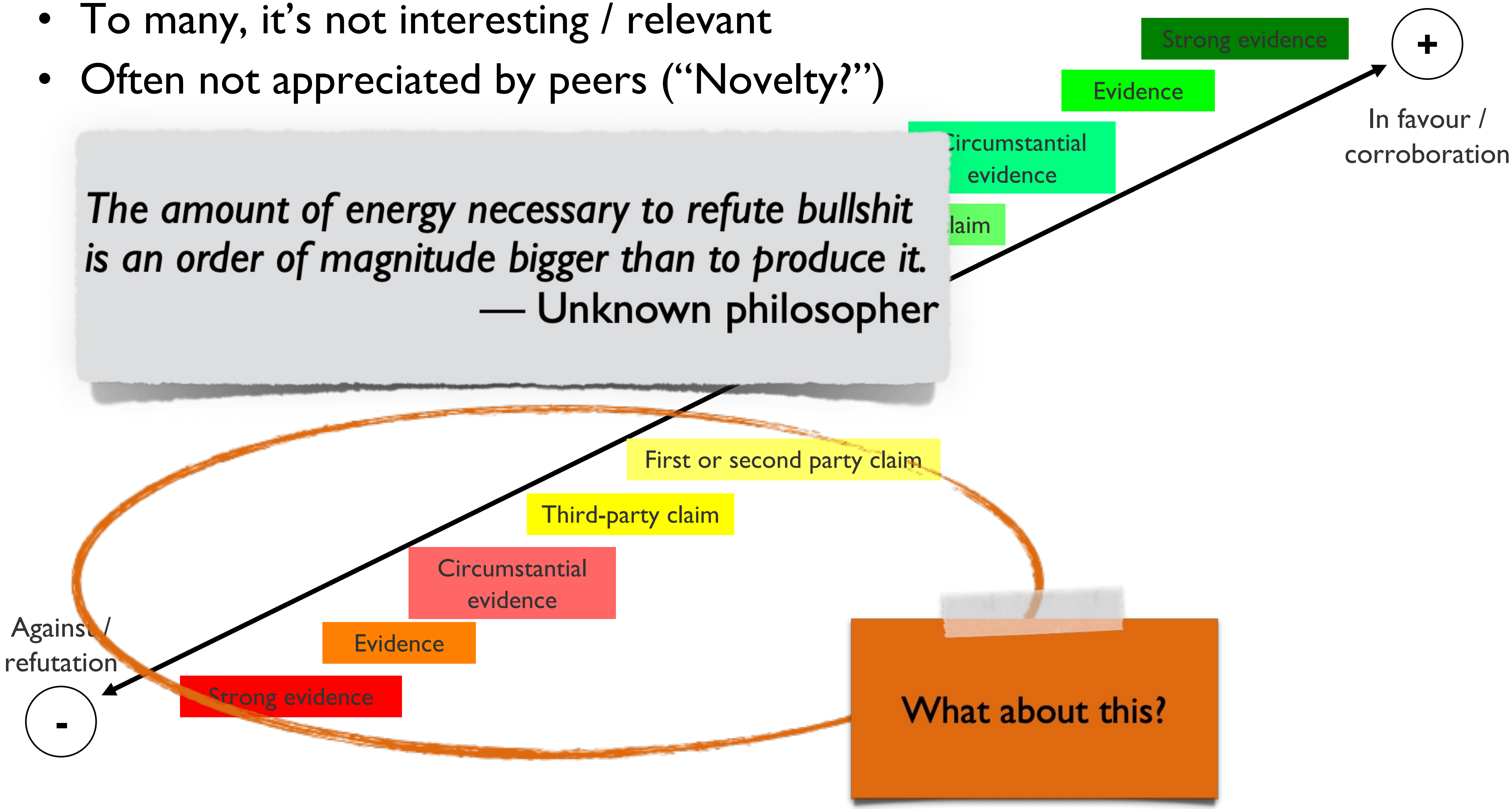
Many reasons for their existence

- Emerged from times where claims by authorities where treated as facts
- Lack of empirical awareness
- Authors do not cite properly
  - Citing claims of (over-)conclusions as facts
  - Citing without reading properly (laziness or no access because work is paywalled)
  - Citing only one side of an argument
  - …



Why not simply debunk (i.e. falsify) folklore?

- It is difficult and very time-consuming
- To many, it's not interesting / relevant
- Often not appreciated by peers ("Novelty?")



> *The amount of energy necessary to refute bullshit is an order of magnitude bigger than to produce it.*
> — Unknown philosopher

Source [for levels of evidence]: Wohlin. An Evidence Profile for Software Engineering Research and Practice, 2013.

Consequences

Limited problem-driven research

- Based (often) on false claims/beliefs
- Little practical/theoretical relevance

Inefficient practice

- Lack of sufficient knowledge
- Lack of efficient methods and tools

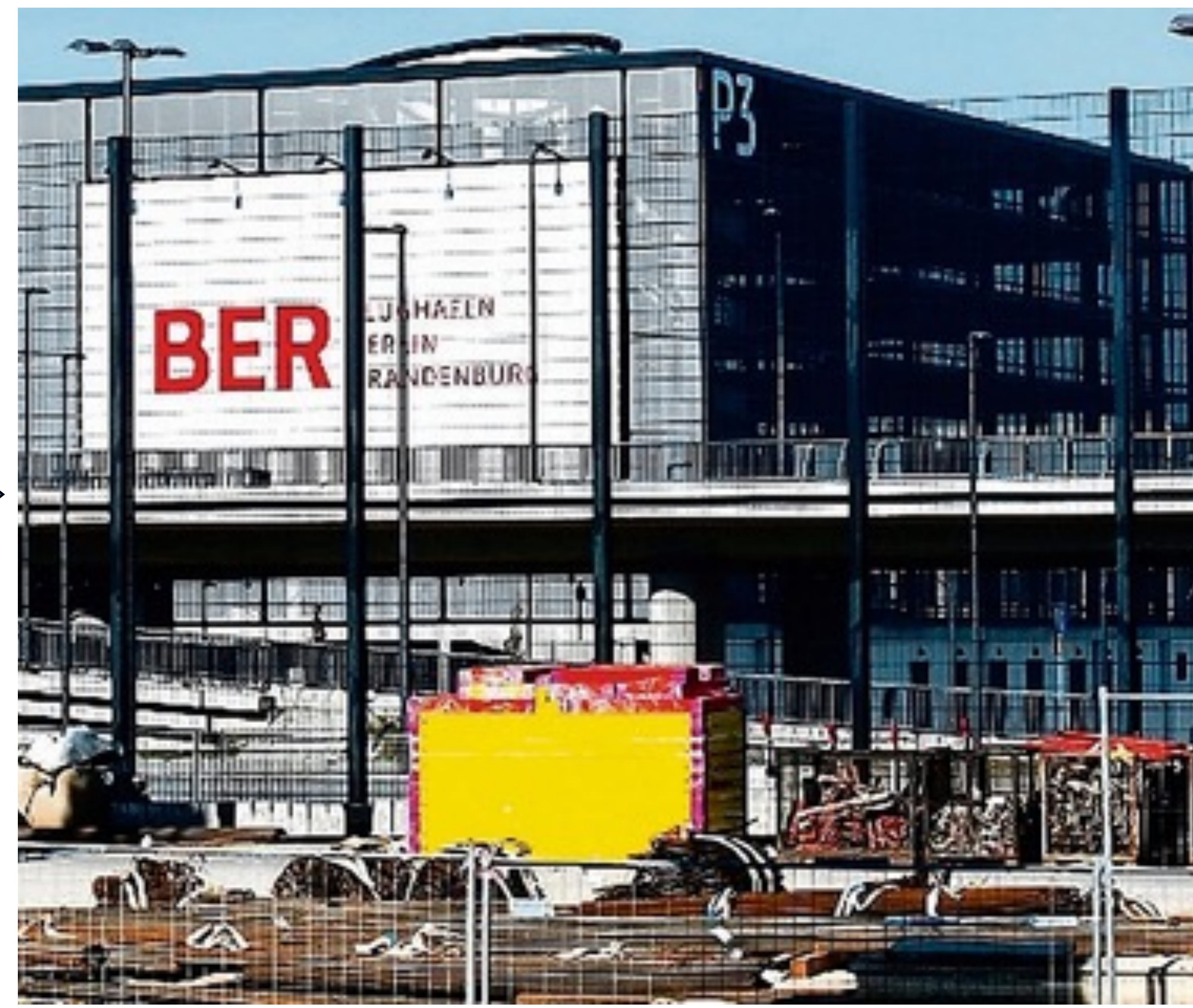


Image Source (l) http://andrewboynton.com/wp-content/uploads/2011/03/IvoryTower.jpg
Image Source (r) http://www.tagesspiegel.de/images/aktueller-stand-am-hauptstadtflughafen/13424442/2-format6001.jpg

Theory building and theory evaluation is crucial in SE

… otherwise, we are not the experimental counterpart
to theoretical computer science, but the homeopathic one.

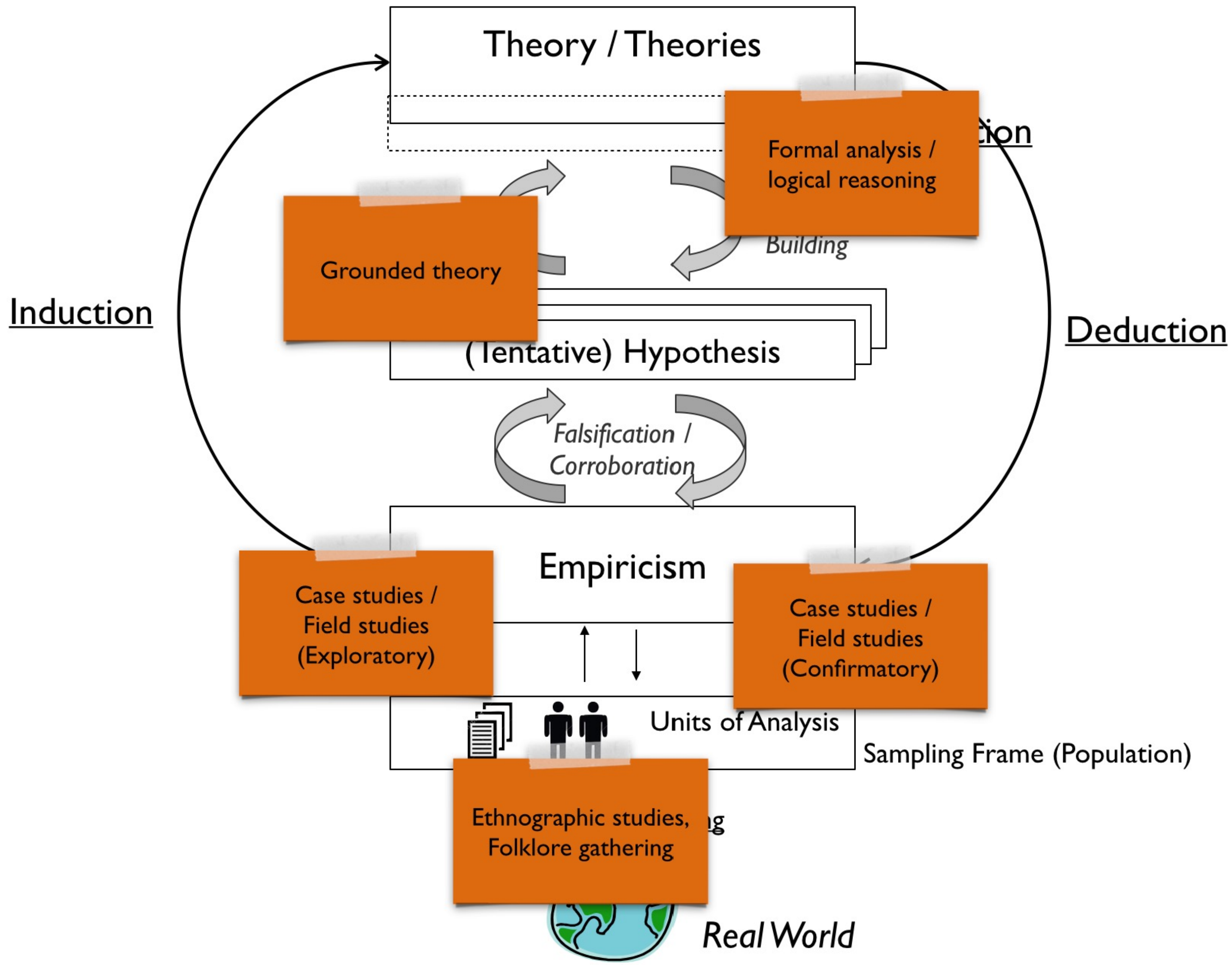
Outline

- Science (in a Nutshell)
- Philosophy of Science - a Historical Perspective
- Key Take Aways
- From Philosophy of Science to Empirical Software Engineering
- **Empirical Software Engineering Processes**
- Current Challenges in Empirical Software Engineering

The **ultimate goal of empirical Software Engineering** processes is **theory building and evaluation** to strengthen and advance our body of knowledge.

But how?

(Reminder)
Progress comes in an iterative, step-wise manner



Each **step** has a specific **objective and purpose.**

Example!

Naming the Pain in Requirements Engineering: Design of a Global Family of Surveys and First Results from Germany

Daniel Méndez Fernández
Technische Universität München
Garching bei München, Germany
mendezfe@in.tum.de

Stefan Wagner
University of Stuttgart
Stuttgart, Germany
stefan.wagner@informatik.uni-stuttgart.de

ABSTRACT

Context: For many years, we have observed industry struggling in defining a high quality requirements engineering (RE) and researchers trying to understand industrial expectations and problems. Although we are investigating the discipline with a plethora of empirical studies, those studies either concentrate on validating specific methods or on single companies or countries. Therefore, they allow only for limited empirical generalisations. **Objective:** To lay an empirical and generalisable foundation about the state of the practice in RE, we aim at a series of open and reproducible surveys that allow us to steer future research in a problem-driven manner. **Method:** We designed a globally distributed family of surveys in joint collaborations with different researchers from different countries. The instrument is based on an initial theory inferred from available studies. As a long-term goal, the survey will be regularly replicated to manifest a clear understanding on the status quo and practical needs in RE. In this paper, we present the design of the family of surveys and first results of its start in Germany. **Results:** Our first results contain responses from 30 German companies. The results are not yet generalisable, but already indicate several trends and problems. For instance, a commonly stated problem respondents see in their company standards are artefacts being underrepresented, and important problems they experience in their projects are incomplete and inconsistent requirements. **Conclusion:** The results suggest that the survey design and instrument are well-suited to be replicated and, thereby, to create a generalisable empirical basis of RE in practice.

Categories and Subject Descriptors

D.2.1 [**Software Engineering**]: Requirements/Specification

General Terms

Experimentation

Keywords

Survey Research, Requirements Engineering, Family of Studies

1. INTRODUCTION

Requirements engineering (RE) is a discipline that constitutes a holistic key to successful development projects as the elicitation, specification and validation of precise and stakeholder-appropriate requirements are critical determinants of quality [2]. At the same time, RE is characterised by the involvement of interdisciplinary stakeholders and uncertainty as many things are not clear from the beginning of a project. Hence, RE is highly volatile and inherently complex by nature.

Although the importance of a high quality RE and its continuos improvement has been recognised for many years, we can still observe industry struggling in defining and applying a high quality RE [14]. The diversity of how RE is performed in various industrial environments, each having its particularities in the domains of application or the software process models used, dooms the discipline to be not only a process area difficult to improve, but also difficult to investigate for common practices and shortcomings.

From a researcher's perspective, experimental research in RE thereby becomes a crucial and challenging task. It is crucial, as experimentation of any kind in RE, ranging from classical action research through observational studies to broad exploratory surveys, are fundamentally necessary to understand the practical needs and improvement goals in RE, to steer problem-driven research and to investigate the value of new RE methods via validation research [4]. It is challenging, because we still need a solid empirical basis that allows for generalisations taking into account the human factors that influence the anyway hardly standardisable discipline like no other in software engineering. In consequence, qualitative research methods are gaining much attention [17], and survey research has become an indispensable means to investigate RE.

1.1 Problem Statement

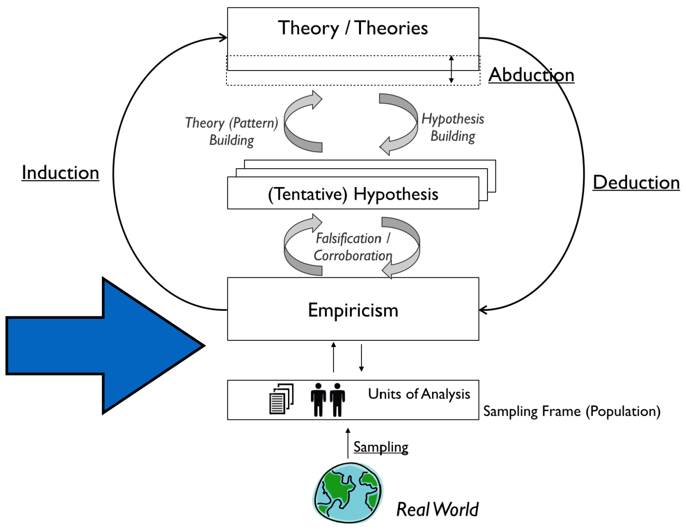
Although we are confident about the value of survey research to understand practical needs and to distill improve-

Research objective / Purpose

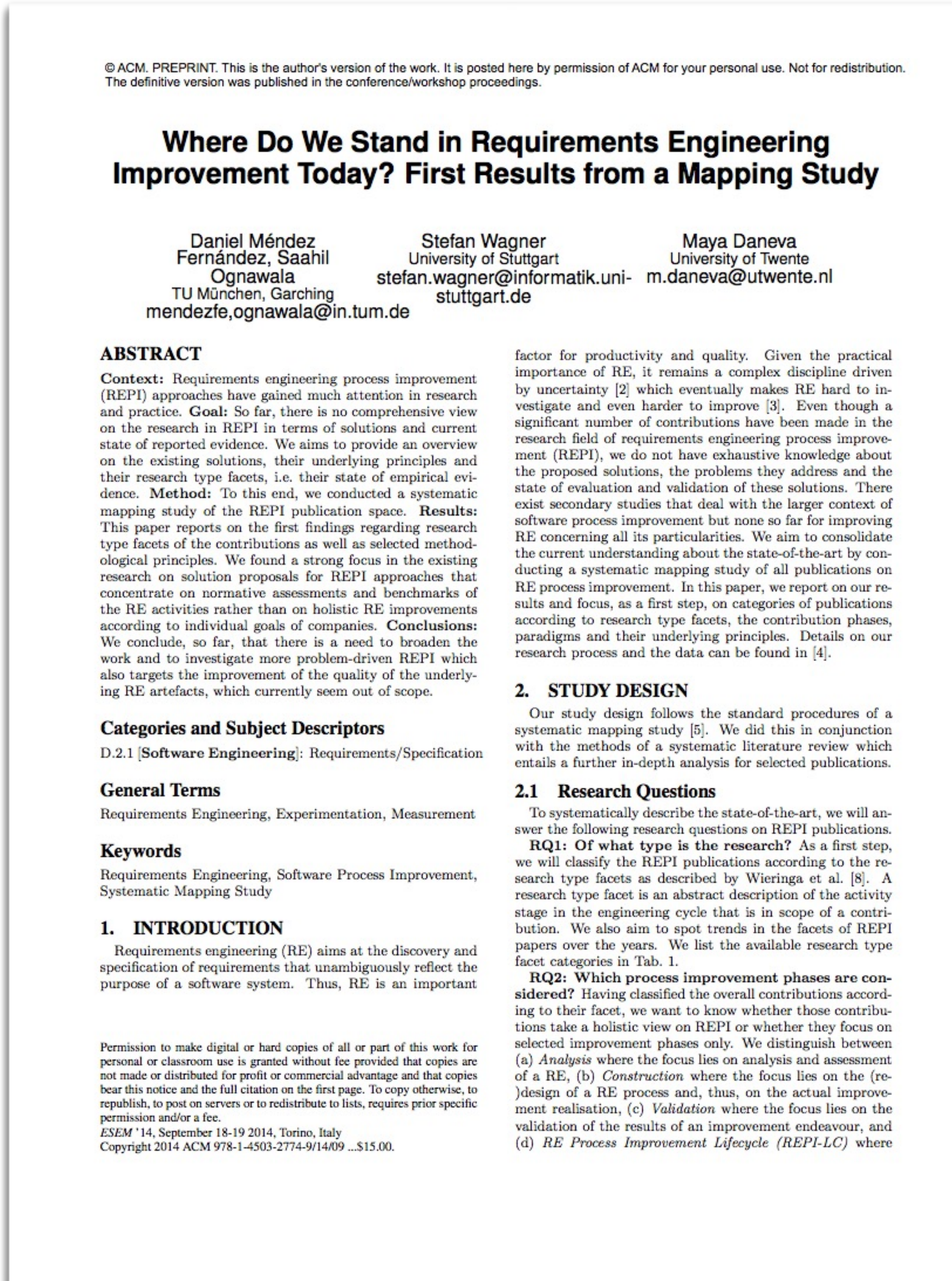
- Exploratory survey to better understand current state of practice and related problems in Requirements Engineering

Method

- (Online) survey research



Example!



Where Do We Stand in Requirements Engineering Improvement Today? First Results from a Mapping Study

Daniel Méndez Fernández, Saahil Ognawala
TU München, Garching
mendezfe,ognawala@in.tum.de

Stefan Wagner
University of Stuttgart
stefan.wagner@informatik.uni-stuttgart.de

Maya Daneva
University of Twente
m.daneva@utwente.nl

ABSTRACT

Context: Requirements engineering process improvement (REPI) approaches have gained much attention in research and practice. **Goal:** So far, there is no comprehensive view on the research in REPI in terms of solutions and current state of reported evidence. We aims to provide an overview on the existing solutions, their underlying principles and their research type facets, i.e. their state of empirical evidence. **Method:** To this end, we conducted a systematic mapping study of the REPI publication space. **Results:** This paper reports on the first findings regarding research type facets of the contributions as well as selected methodological principles. We found a strong focus in the existing research on solution proposals for REPI approaches that concentrate on normative assessments and benchmarks of the RE activities rather than on holistic RE improvements according to individual goals of companies. **Conclusions:** We conclude, so far, that there is a need to broaden the work and to investigate more problem-driven REPI which also targets the improvement of the quality of the underlying RE artefacts, which currently seem out of scope.

Categories and Subject Descriptors

D.2.1 [**Software Engineering**]: Requirements/Specification

General Terms

Requirements Engineering, Experimentation, Measurement

Keywords

Requirements Engineering, Software Process Improvement, Systematic Mapping Study

1. INTRODUCTION

Requirements engineering (RE) aims at the discovery and specification of requirements that unambiguously reflect the purpose of a software system. Thus, RE is an important factor for productivity and quality. Given the practical importance of RE, it remains a complex discipline driven by uncertainty [2] which eventually makes RE hard to investigate and even harder to improve [3]. Even though a significant number of contributions have been made in the research field of requirements engineering process improvement (REPI), we do not have exhaustive knowledge about the proposed solutions, the problems they address and the state of evaluation and validation of these solutions. There exist secondary studies that deal with the larger context of software process improvement but none so far for improving RE concerning all its particularities. We aim to consolidate the current understanding about the state-of-the-art by conducting a systematic mapping study of all publications on RE process improvement. In this paper, we report on our results and focus, as a first step, on categories of publications according to research type facets, the contribution phases, paradigms and their underlying principles. Details on our research process and the data can be found in [4].

2. STUDY DESIGN

Our study design follows the standard procedures of a systematic mapping study [5]. We did this in conjunction with the methods of a systematic literature review which entails a further in-depth analysis for selected publications.

2.1 Research Questions

To systematically describe the state-of-the-art, we will answer the following research questions on REPI publications.

RQ1: Of what type is the research? As a first step, we will classify the REPI publications according to the research type facets as described by Wieringa et al. [8]. A research type facet is an abstract description of the activity stage in the engineering cycle that is in scope of a contribution. We also aim to spot trends in the facets of REPI papers over the years. We list the available research type facet categories in Tab. 1.

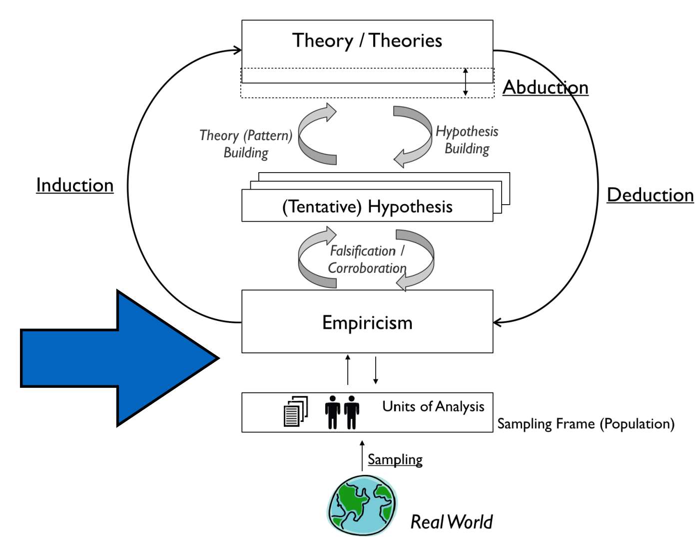
RQ2: Which process improvement phases are considered? Having classified the overall contributions according to their facet, we want to know whether those contributions take a holistic view on REPI or whether they focus on selected improvement phases only. We distinguish between (a) *Analysis* where the focus lies on analysis and assessment of a RE, (b) *Construction* where the focus lies on the (re-)design of a RE process and, thus, on the actual improvement realisation, (c) *Validation* where the focus lies on the validation of the results of an improvement endeavour, and (d) *RE Process Improvement Lifecycle (REPI-LC)* where

Research objective / Purpose

- Exploratory literature study to understand current state of reported evidence in Requirements Engineering (process) improvement and potential gaps

Method

- Systematic mapping study



Example!

Improving Requirements Engineering by Artefact Orientation

Daniel Méndez Fernández[1] and Roel Wieringa[2]

[1] Technische Universität München, Germany
http://www4.in.tum.de/~mendezfe
[2] University of Twente, Netherlands
http://www.cs.utwente.nl/~roelw

Abstract. The importance of continuously improving requirements engineering (RE) has been recognised for many years. Similar to available software process improvement approaches, most RE improvement approaches focus on a normative and solution-driven assessment of companies rather than on a problem-driven RE improvement. The approaches dictate the implementation of a one-size-fits-all reference model without doing a proper problem investigation first, whereas the notion of quality factually depends on whether RE achieves company-specific goals. The approaches furthermore propagate process areas and methods, without proper awareness of the quality in the created artefacts on which the quality of many development phases rely. Little knowledge exists about how to conduct a problem-driven RE improvement that gives attention to the improvement of the artefacts. A promising solution is to start an improvement with an empirical investigation of the RE stakeholders, goals, and artefacts in the company to identify problems while abstracting from inherently complex processes. The RE improvement is then defined and implemented in joint action research workshops with the stakeholders to validate potential solutions while again concentrating on the artefacts. In this paper, we contribute an artefact-based, problem-driven RE improvement approach that emerged from a series of completed RE improvements. We discuss lessons learnt and present first result from an ongoing empirical evaluation at a German company. Our results suggest that our approach supports process engineers in a problem-driven RE improvement, but we need deeper examination of the resulting RE company standard, which is in scope of the final evaluation.

Keywords: Requirements Engineering, Artefact Orientation, Empirical Design Science, Software Process Improvement

1 Introduction

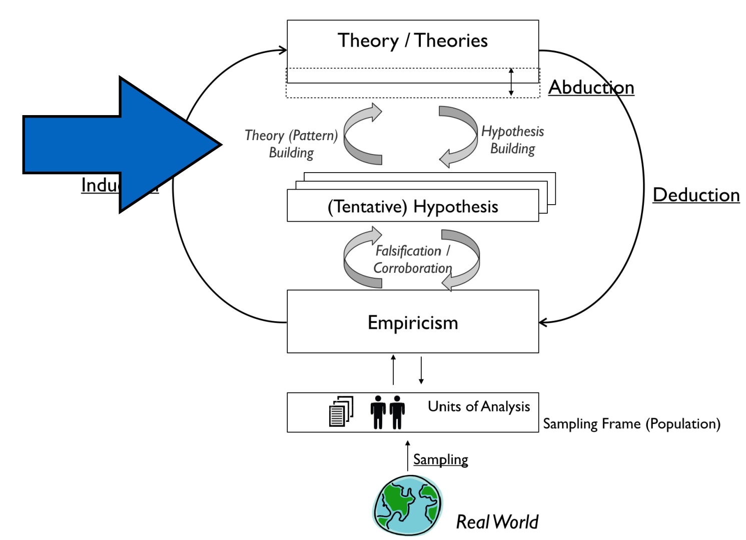
Requirements engineering (RE) constitutes an important success factor for software development projects, since stakeholder-appropriate requirements are important determinants of quality. Incorrect or missing requirements can greatly add to the implementation or maintenance effort later. At the same time, RE is an interdisciplinary area in a software development process that is driven by

Research objective / Purpose

- Design of an RE improvement approach by synthesising existing concepts

Method

- (Design) theory building



Example!

A Case Study on
Artefact-based RE Improvement in Practice

Daniel Méndez Fernández[1] and Stefan Wagner[2]

[1] Technische Universität München, Germany
http://www4.in.tum.de/~mendezfe
[2] University of Stuttgart, Germany
http://www.iste.uni-stuttgart.de/

Abstract. *Background:* Most requirements engineering (RE) process improvement approaches are solution-driven and activity-based. They focus on the assessment of the RE of a company against an external norm of best practices. A consequence is that practitioners often have to rely on an improvement approach that skips a profound problem analysis and that results in an RE approach that might be alien to the organisational needs. *Objective:* In recent years, we have developed an RE improvement approach (called *ArtREPI*) that guides a holistic RE improvement against individual goals of a company putting primary attention to the quality of the artefacts. In this paper, we aim at exploring ArtREPI's benefits and limitations. *Method:* We contribute an industrial evaluation of ArtREPI by relying on a case study research. *Results:* Our results suggest that ArtREPI is well-suited for the establishment of an RE that reflects a specific organisational culture but to some extent at the cost of efficiency resulting from intensive discussions on a terminology that suits all involved stakeholders. *Conclusions:* Our results reveal first benefits and limitations, but we can also conclude the need of longitudinal and independent investigations for which we herewith lay the foundation.

Keywords: Requirements Engineering, Artefact Orientation, Software Process Improvement, Case Study Research

1 Introduction

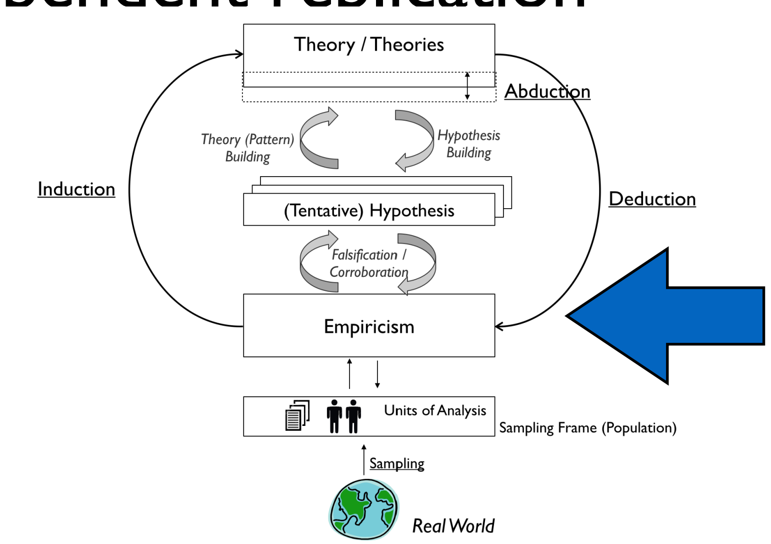
Requirements engineering (RE) constitutes an important success factor for software development projects since stakeholder-appropriate requirements are important determinants of quality. Its interdisciplinary nature, the uncertainty, and the complexity in the process, however, make the discipline difficult to investigate and to improve [1]. For an RE improvement, process engineers have to decide whether to opt for *problem orientation* or for *solution orientation* [2,3]. In a solution-driven improvement, the engineers assess and adapt their RE reference model, which provides a company-specific blueprint of RE practices and artefacts, against an external norm of best practices. The latter is meant to lead to a high quality RE based on universal, external goals (see, e.g. CMMI for

Research objective / Purpose

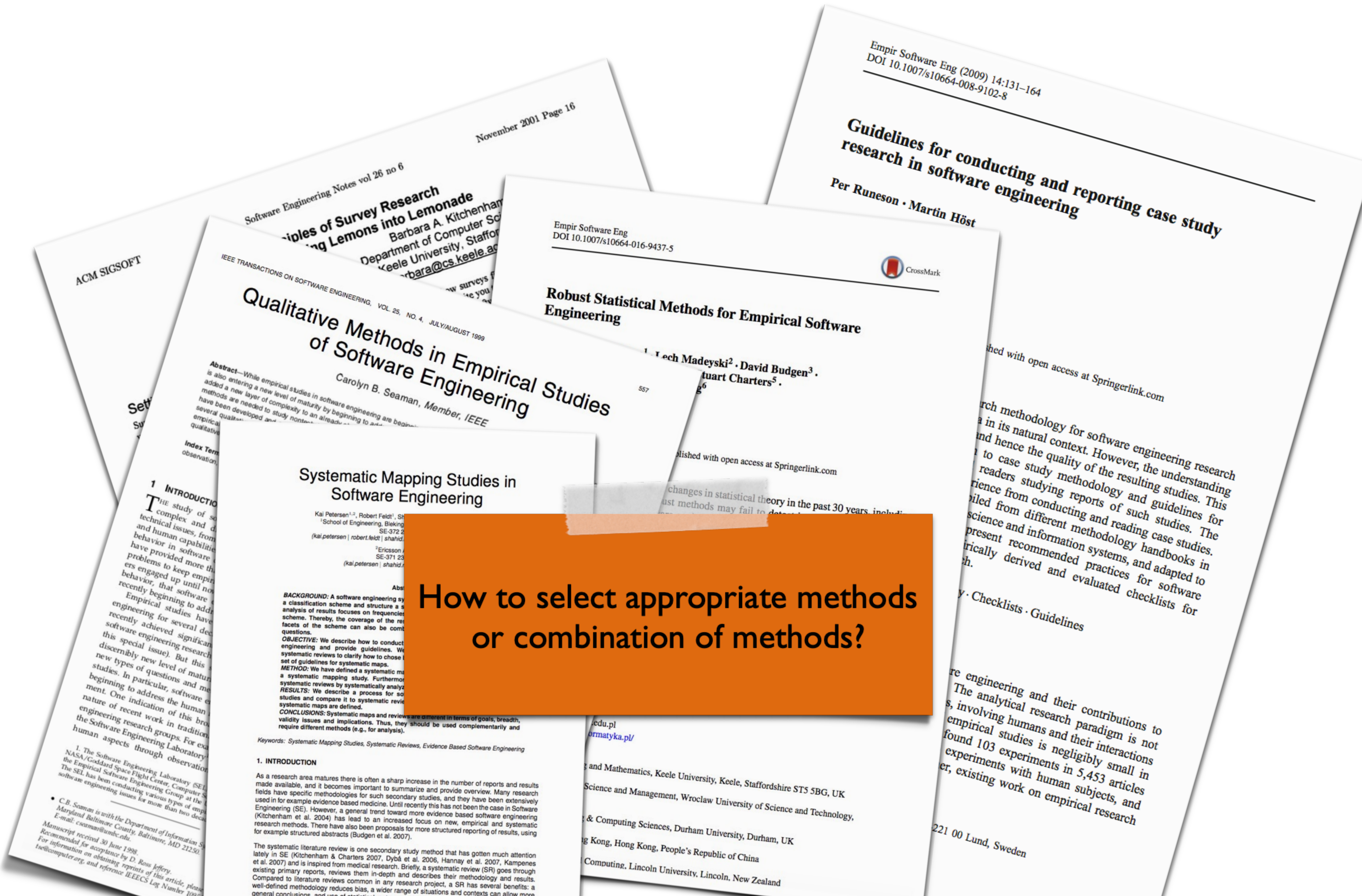
- Comparative case study to understand benefits and limitations when improving RE following a specific approach

Method

- Case study research with canonical action research
- Independent replication



Different objectives require different methods



There is no universal silver-bullet*

* Reminder: No universal way of scientific working.

Empirical processes: an abstract view

Planning and Definition	• Identify and outline problem (area) • Determine research objectives
Method and Strategy Selection	• Select type of study and method(s) • Identify necessary environment (including units of analysis)
Design and (Method) Execution	• Design and validate study protocol (and validity procedures) • Employ research method following respective (detailed) processes
Conclusion Drawing	• Analyse data • Reflect on potential threats to validity
Packaging and Reporting	• Package (and ideally disclose) data • Report on results (in tune with audience)

Empirical processes: an abstract view

- Planning and Definition
- Method and Strategy Selection
- Design and (Method) Execution
- Conclusion Drawing
- Packaging and Reporting

Which method(s) do we need to employ?

Scope of detailed empirical methods

Planning and definition

Planning and Definition

Method and Strategy Selection

Design and (Method) Execution

Conclusion Drawing

Packaging and Reporting

At the end of the planning phase, we need to know:

- Why should the empirical study be conducted (purpose and goal)?
- What will be investigated?

Steps to get there:

- Identify (potential) problems
- Select problem in scope of study
- Formulate research goal / questions

Problem identification

By the way
The problem identification can comprehend own (or even multiple) studies

What is the goal?

- Identify open (theoretical and / or practical) problems

What could be good starting points?

- Existing (i.e. reported) hypotheses or theories
- Claims or assumptions about, e.g., a technologies effectiveness
- Results that contradict common hypotheses or theories

Analyse the state of the art

- (Systematic) literature reviews / mapping studies

(complementarily) Analyse the state of the practice

- Document analysis (projects, public repositories, etc.)
- Interviews, surveys, and observations

Problem selection

Scientific criteria

- How does its investigation contribute to research (theoretical relevance)?
- To which extent can it be investigated empirically?
- …

Practical criteria

- To which extent is it a practical problem (practical relevance)?
- To which extent does the problem depend on particularities of a practical context?
- …

Ethical (and also pragmatic) criteria

- Does the investigation imply (personal) benefits, disadvantages, risks, harms?
- Is it necessary and possible to collect and keep data anonymous?
- How could and should the results (and data) be published?
- …

Type of research goals (and purposes of methodologies)

	Explanatory	Exploratory	Descriptive	Improving
Scope	• Seeking an explanation of a situation or a problem • Mostly but not necessarily in the form of a causal relation-ship	• Finding out what is happening, seeking new insights and generating ideas and hypotheses for new research • Understanding events, decisions, processes, …, and their meaning in specific context based on subjects' insight	• Portraying a situation or phenomenon. • Drawing accurate descriptions of events, decisions, processes,… , and the relations among them	• Trying to improve a certain aspect of the studied phenomenon **Prerequisites** • Baseline models (practice) • Standards • Oracles
Basis for…	• … precise hypothesis and theories • …prediction models	• …new (tentative and vague) hypothesis (out of curiosity-driven research)	• … precise hypothesis and theories	• … under-standing the impact of artefacts

Based on: Runeson, P, Höst, M. Guidelines for conducting and reporting case study research in software engineering, 2009.

Research goal definition

A clearly structured goal supports the reproducibility of a research endeavour!

Analyse ______________________________

(units of analysis: process, product, people, …)

for the purpose of ______________________________

(purpose: understand, describe, explain, evaluate, change, …)

with respect to ______________________________

(quality focus: cost, correctness, reliability, usability, ...)

from the point of view of ______________________________

(stakeholder: user, customer, manager, developer, corporation,. ..)

in the context ______________________________

(context: problem, people, resource, or process factors, ...)

Based on: Basili, V., Caldiera1, G., Rombach, D.The Goal Question Metric Approach, 1994.

Example!

Research goal definition

Analyse *a problem-driven requirements engineering improvement approach*
(units of analysis: process, product, people, …)

for the purpose of *evaluation*
(purpose: understand, describe, explain, evaluate, change, …)

with respect to *usability (inter alia)*
(quality focus: cost, correctness, reliability, usability, ...)

from the point of view of *(process) engineers*
(stakeholder: user, customer, manager, developer, corporatio

in the context *custom software development projects*
(context: problem, people, resource, or process factors, ...)

A Case Study on Artefact-based RE Improvement in Practice

Daniel Méndez Fernández[1] and Stefan Wagner[2]

[1] Technische Universität München, Germany
http://www4.in.tum.de/~mendezfe
[2] University of Stuttgart, Germany
http://www.iste.uni-stuttgart.de/

Abstract. *Background:* Most requirements engineering (RE) process improvement approaches are solution-driven and activity-based. They focus on the assessment of the RE of a company against an external norm of best practices. A consequence is that practitioners often have to rely on an improvement approach that skips a profound problem analysis and that results in an RE approach that might be alien to the organisational needs. *Objective:* In recent years, we have developed an RE improvement approach (called *ArtREPI*) that guides a holistic RE improvement against individual goals of a company putting primary attention to the quality of the artefacts. In this paper, we aim at exploring ArtREPI's benefits and limitations. *Method:* We contribute an industrial evaluation of ArtREPI by relying on a case study research. *Results:* Our results suggest that ArtREPI is well-suited for the establishment of an RE that reflects a specific organisational culture but to some extent at the cost of efficiency resulting from intensive discussions on a terminology that suits all involved stakeholders. *Conclusions:* Our results reveal first benefits and limitations, but we can also conclude the need of longitudinal and independent investigations for which we herewith lay the foundation.

Keywords: Requirements Engineering, Artefact Orientation, Software Process Improvement, Case Study Research

1 Introduction

Requirements engineering (RE) constitutes an important success factor for software development projects since stakeholder-appropriate requirements are important determinants of quality. Its interdisciplinary nature, the uncertainty, and the complexity in the process, however, make the discipline difficult to investigate and to improve [1]. For an RE improvement, process engineers have to decide whether to opt for *problem orientation* or for *solution orientation* [2,3]. In a solution-driven improvement, the engineers assess and adapt their RE reference model, which provides a company-specific blueprint of RE practices and artefacts, against an external norm of best practices. The latter is meant to lead to a high quality RE based on universal, external goals (see, e.g. CMMI for

From research goals to research questions

Non-causal research questions

- What is X? What does X mean?
- What are the differences between X1 and X2?
- How does X work? Why / why not?
- How do you select/adopt/use/estimate/…. X?
- Why does a subject support/select/adopt/use/…. X?

Casual research question

- Does X cause Y?
- Does X1 cause more of Y than X2 causes of Y?

From research goals to research questions

Example!

Non-causal research questions

RQ 1 How well are process engineers supported in their RE improvement tasks?

RQ 2 How well are project participants supported by the resulting RE reference model?

A Case Study on
Artefact-based RE Improvement in Practice

Daniel Méndez Fernández[1] and Stefan Wagner[2]

[1] Technische Universität München, Germany
http://www4.in.tum.de/~mendezfe
[2] University of Stuttgart, Germany
http://www.iste.uni-stuttgart.de/

Abstract. *Background:* Most requirements engineering (RE) process improvement approaches are solution-driven and activity-based. They focus on the assessment of the RE of a company against an external norm of best practices. A consequence is that practitioners often have to rely on an improvement approach that skips a profound problem analysis and that results in an RE approach that might be alien to the organisational needs. *Objective:* In recent years, we have developed an RE improvement approach (called *ArtREPI*) that guides a holistic RE improvement against individual goals of a company putting primary attention to the quality of the artefacts. In this paper, we aim at exploring ArtREPI's benefits and limitations. *Method:* We contribute an industrial evaluation of ArtREPI by relying on a case study research. *Results:* Our results suggest that ArtREPI is well-suited for the establishment of an RE that reflects a specific organisational culture but to some extent at the cost of efficiency resulting from intensive discussions on a terminology that suits all involved stakeholders. *Conclusions:* Our results reveal first benefits and limitations, but we can also conclude the need of longitudinal and independent investigations for which we herewith lay the foundation.

Keywords: Requirements Engineering, Artefact Orientation, Software Process Improvement, Case Study Research

1 Introduction

Requirements engineering (RE) constitutes an important success factor for software development projects since stakeholder-appropriate requirements are important determinants of quality. Its interdisciplinary nature, the uncertainty, and the complexity in the process, however, make the discipline difficult to investigate and to improve [1]. For an RE improvement, process engineers have to decide whether to opt for *problem orientation* or for *solution orientation* [2,3]. In a solution-driven improvement, the engineers assess and adapt their RE reference model, which provides a company-specific blueprint of RE practices and artefacts, against an external norm of best practices. The latter is meant to lead to a high quality RE based on universal, external goals (see, e.g. CMMI for

Method and strategy selection

- Planning and Definition
- **Method and Strategy Selection**
- Design and (Method) Execution
- Conclusion Drawing
- Packaging and Reporting

At the end of the method selection phase, we need to know:

- What type of study do we need to conduct?
- Which empirical method(s) do we need?
- What is the necessary environment?

Steps to get there:

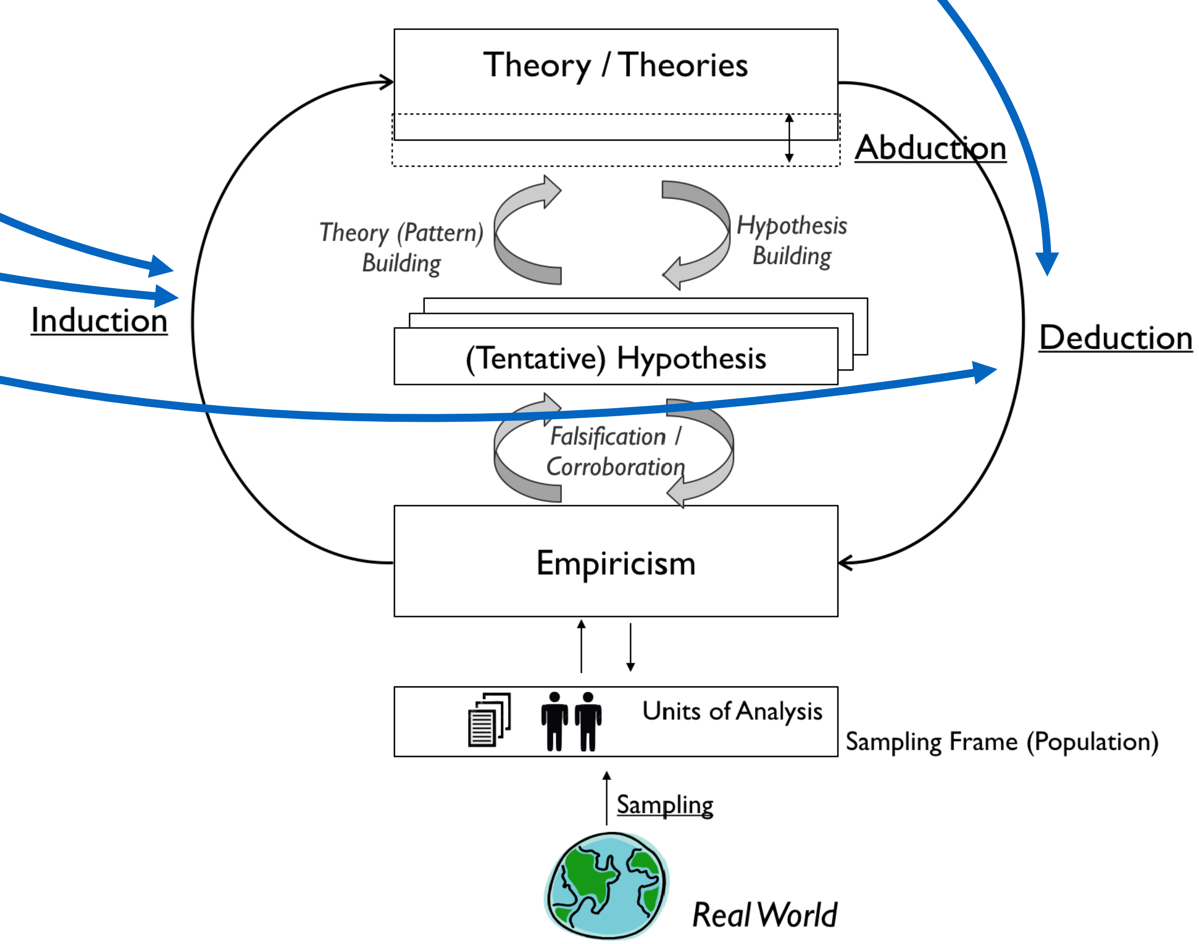
- Identify method(s) and environment based on goals and purpose
- Reflect on further important decision criteria (often coming with a trade-off)

What is the nature of the study? (Inductive? Deductive? Both?)

	Explanatory	Exploratory	Descriptive	Improving
Scope	• Seeking an explanation of a situation or a problem • Mostly but not necessarily in the form of a causal relationship	• Finding out what is happening, seeking new insights and generating ideas and hypotheses for new research	• Portraying a situation or phenomenon.	• Trying to improve a certain aspect of the studied phenomenon

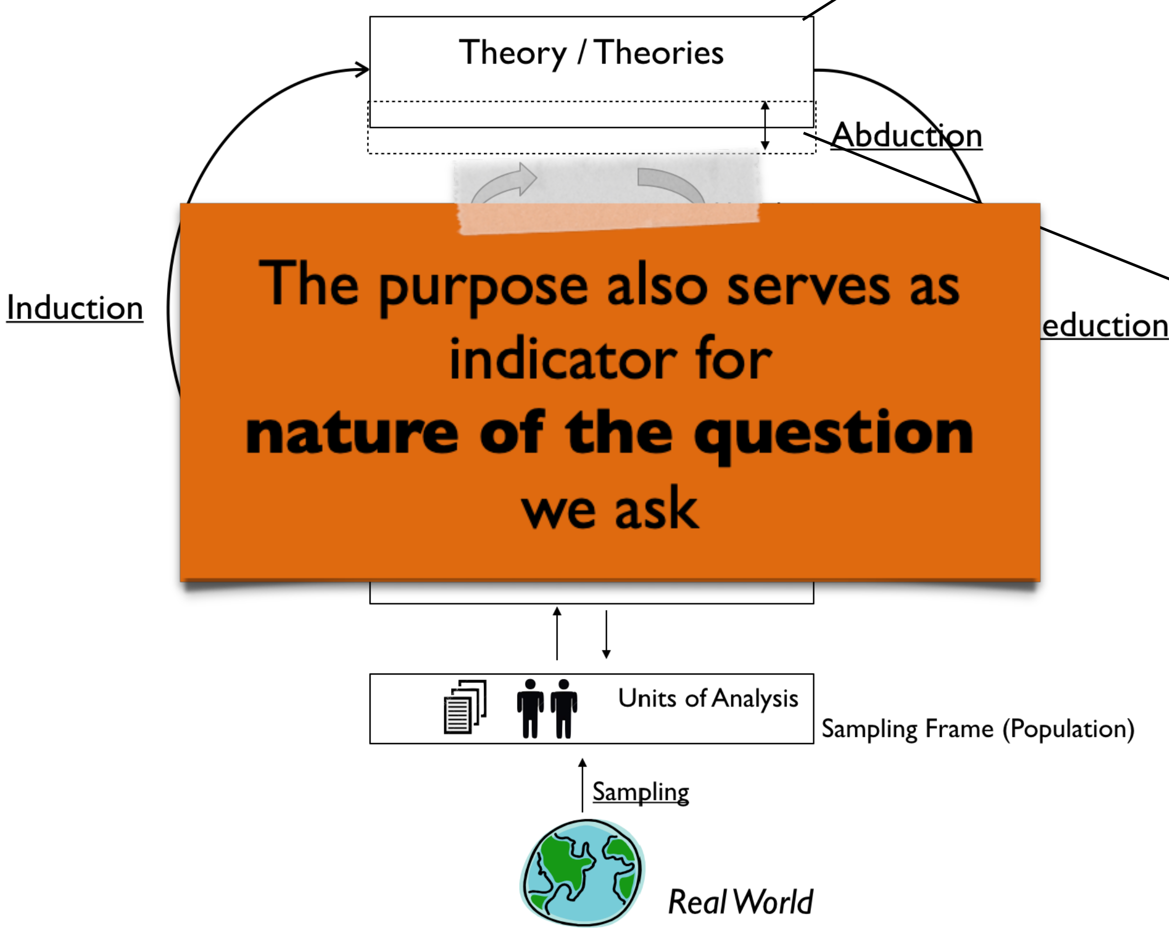
Purpose of methodology

Research goals and study purpose serve as first indicator for the **nature of the study** *



* To be seen as indicators only

What is the relation to the existing body of knowledge?



The purpose also serves as indicator for **nature of the question** we ask

Purpose of theory

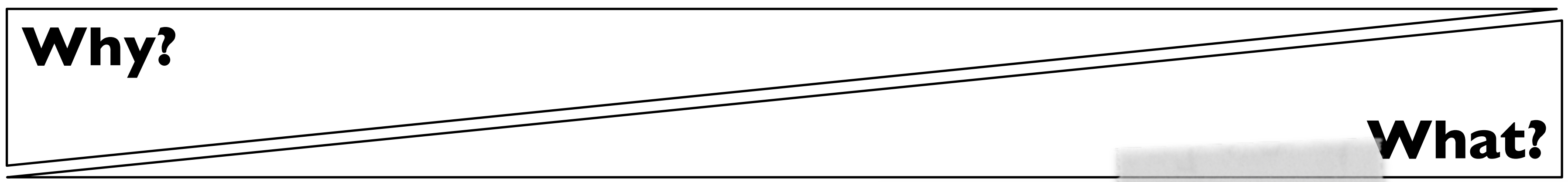
	Analytical	Predictive	Explanatory	Explanatory & Predictive
Scope	• Descriptions and con-ceptualisations, including taxonomies, classifications, and ontologies - What is?	• Prediction of what will happen in the future - What will happen?	• Identification of phenomena by identifying causes, mechanisms or reasons - Why is?	• Prediction of what will happen in the future and explanation - What will happen and why?

- Building a new theory?
- "Testing"/Modifying existing theory?

	Explanatory	Exploratory	Descriptive	Improving
Scope	• Seeking an explanation of a situation or a problem • Mostly but not necessarily in the form of a causal relationship	• Finding out what is happening, seeking new insights and generating ideas and hypotheses for new research	• Portraying a situation or phenomenon.	• Trying to improve a certain aspect of the studied phenomenon

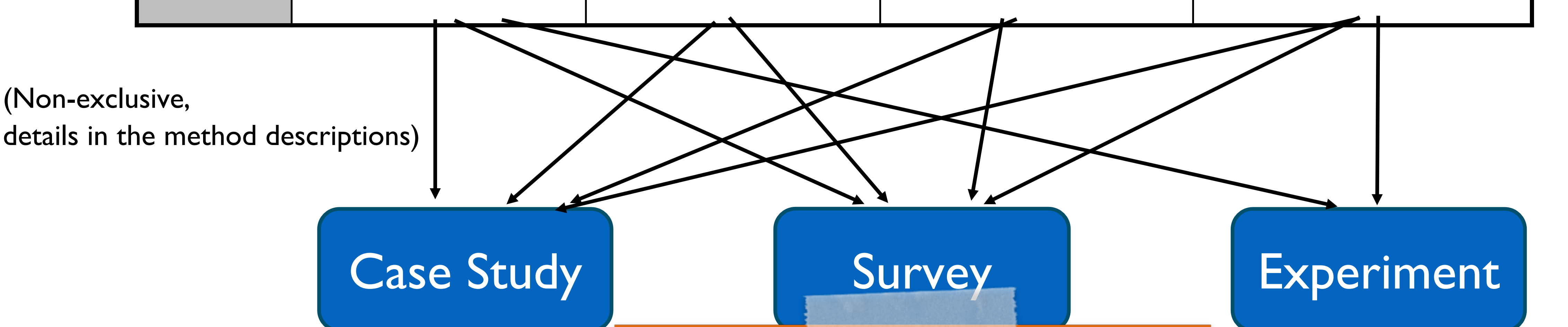
Purpose of methodology

What is the nature of the question we ask? (*What* versus *Why*)



Nature of the question we ask serves as indicator for the **method** we should employ

	Explanatory	Exploratory	De	
Scope	• Seeking an explanation of a situation or a problem • Mostly but not necessarily in the form of a causal relationship	• Finding out what is happening, seeking new insights and generating ideas and hypotheses for new research	• Portray pheno	



The methods serve as indicator for the necessary **environment**

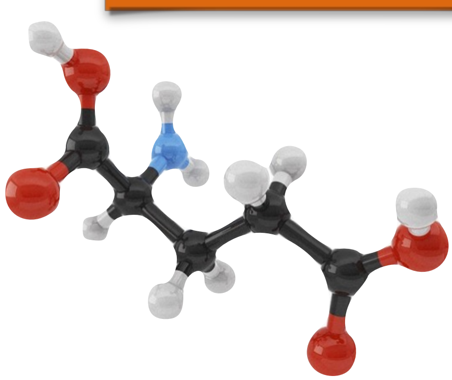
What is the nature of the environment?

What? ⟷ **Why?**

(Need for) reality →

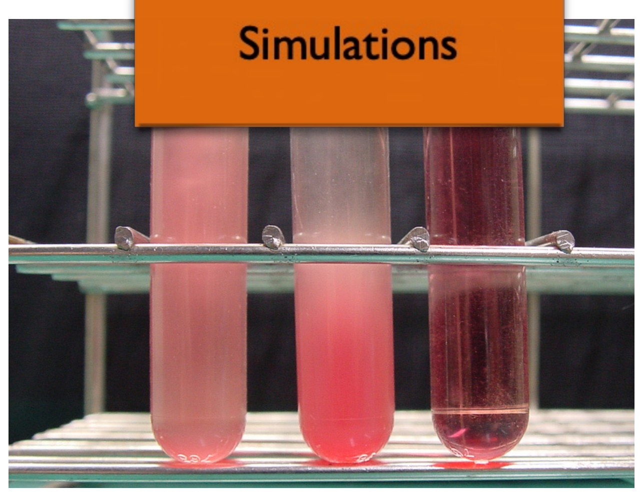
← **(Need for) control**

Formal methods and models



Theoretical space

Simulations



Environment made irrelevant

Controlled experiments

Artificial (controlled) environment

Case study research

Realistic environment

Artificial — *Reality*

What is the population source?

Artificial | Reality

An (imperfect) universe of possibilities

What are the units of analysis?





Academia.edu

GitHub

- Artefacts (e.g. specification documents, log files)
- People (e.g. Java developers, process engineers) or
- Groups of people (e.g. teams, companies)

Method and strategy selection: Summary of important decision criteria

- What is the purpose of the study?
  - Exploratory? Descriptive? Explanatory? Improving?
- What is the nature of the study?
  - Inductive? Deductive?
- What is the relation to the existing body of knowledge?
  - Building a new theory? Testing existing theory?
- What is the nature of the questions we ask?
  - What-questions? Why-questions?
- What is the nature of the environment?
  - Controlled environments? Realistic environments?
- What is the necessary sample?
  - Population source?
  - Units of analysis?

Criteria for selecting methods

Non-exclusive and non-sequential

Criteria for environment selection (and sampling)

Method and strategy selection: Summary of important decision criteria

Example!

- What is the purpose of the study?
  - Improving
- What is the nature of the study?
  - Deductive
- What is the relation to the existing body of knowledge?
  - Testing existing (design) theory
- What is the nature of the questions we ask?
  - Why-questions
- What is the nature of the environment?
  - Realistic environment
- What is the necessary sample?
  - Group of process engineers in a custom software development team

A Case Study on
Artefact-based RE Improvement in Practice

Daniel Méndez Fernández[1] and Stefan Wagner[2]

[1] Technische Universität München, Germany
http://www4.in.tum.de/~mendezfe
[2] University of Stuttgart, Germany
http://www.iste.uni-stuttgart.de/

Abstract. *Background:* Most requirements engineering (RE) process improvement approaches are solution-driven and activity-based. They focus on the assessment of the RE of a company against an external norm of best practices. A consequence is that practitioners often have to rely on an improvement approach that skips a profound problem analysis and that results in an RE approach that might be alien to the organisational needs. *Objective:* In recent years, we have developed an RE improvement approach (called *ArtREPI*) that guides a holistic RE improvement against individual goals of a company putting primary attention to the quality of the artefacts. In this paper, we aim at exploring ArtREPI's benefits and limitations. *Method:* We contribute an industrial evaluation of ArtREPI by relying on a case study research. *Results:* Our results suggest that ArtREPI is well-suited for the establishment of an RE that reflects a specific organisational culture but to some extent at the cost of efficiency resulting from intensive discussions on a terminology that suits all involved stakeholders. *Conclusions:* Our results reveal first benefits and limitations, but we can also conclude the need of longitudinal and independent investigations for which we herewith lay the foundation.

Keywords: Requirements Engineering, Artefact Orientation, Software Process Improvement, Case Study Research

1 Introduction

Requirements engineering (RE) constitutes an important success factor for software development projects since stakeholder-appropriate requirements are important determinants of quality. Its interdisciplinary nature, the uncertainty, and the complexity in the process, however, make the discipline difficult to investigate and to improve [1]. For an RE improvement, process engineers have to decide whether to opt for *problem orientation* or for *solution orientation* [2,3]. In a solution-driven improvement, the engineers assess and adapt their RE reference model, which provides a company-specific blueprint of RE practices and artefacts, against an external norm of best practices. The latter is meant to lead to a high quality RE based on universal, external goals (see, e.g. CMMI for

Is that all?

No (of course not).

Scientific working is influenced by various criteria

- Purpose of the methodology
- Degree of realism and control
- Scope of the study (and validity)
- Theoretical impact
- Practical impact
- Usefulness of emerging theories to researchers and practitioners
- Access to data
- Risk of failure
- Time and cost
- …

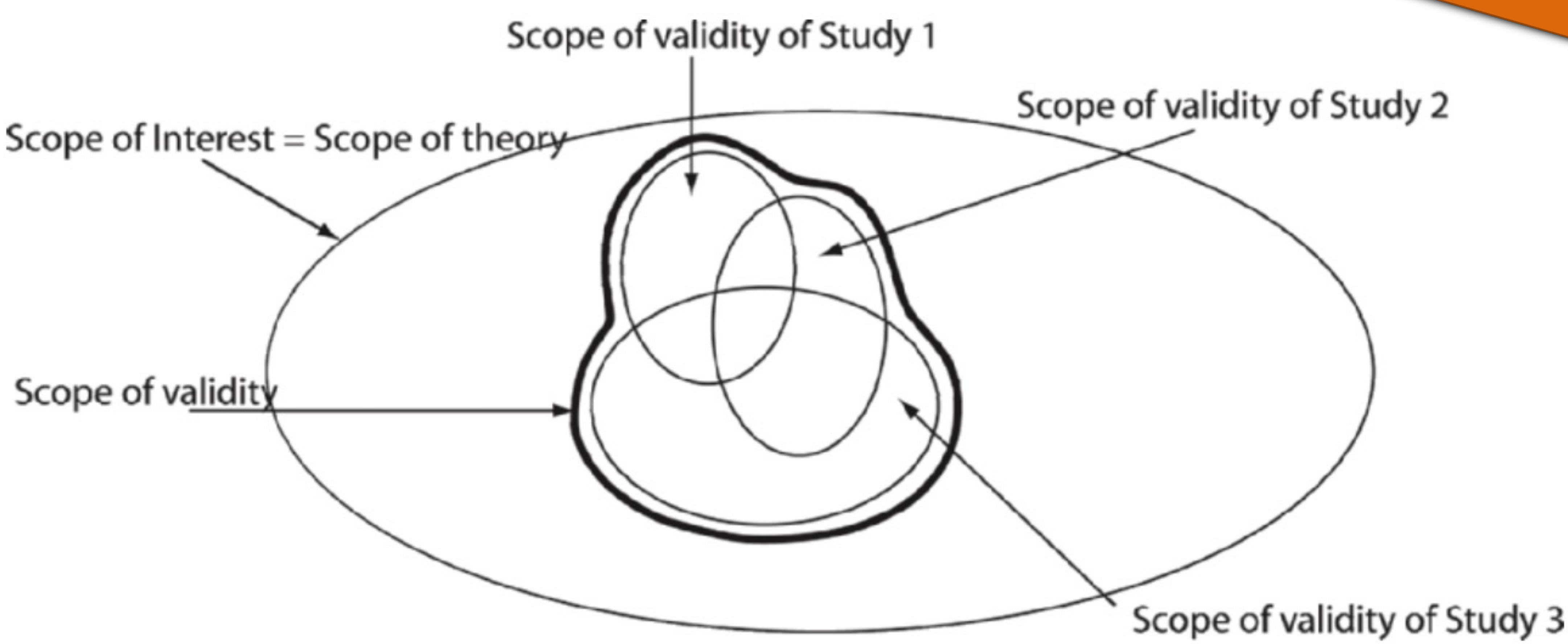
Reminder
Scientific working is also influenced by social and economic aspects (trade-offs)

How do we achieve scientific progress?

In an iterative and step-wise manner.

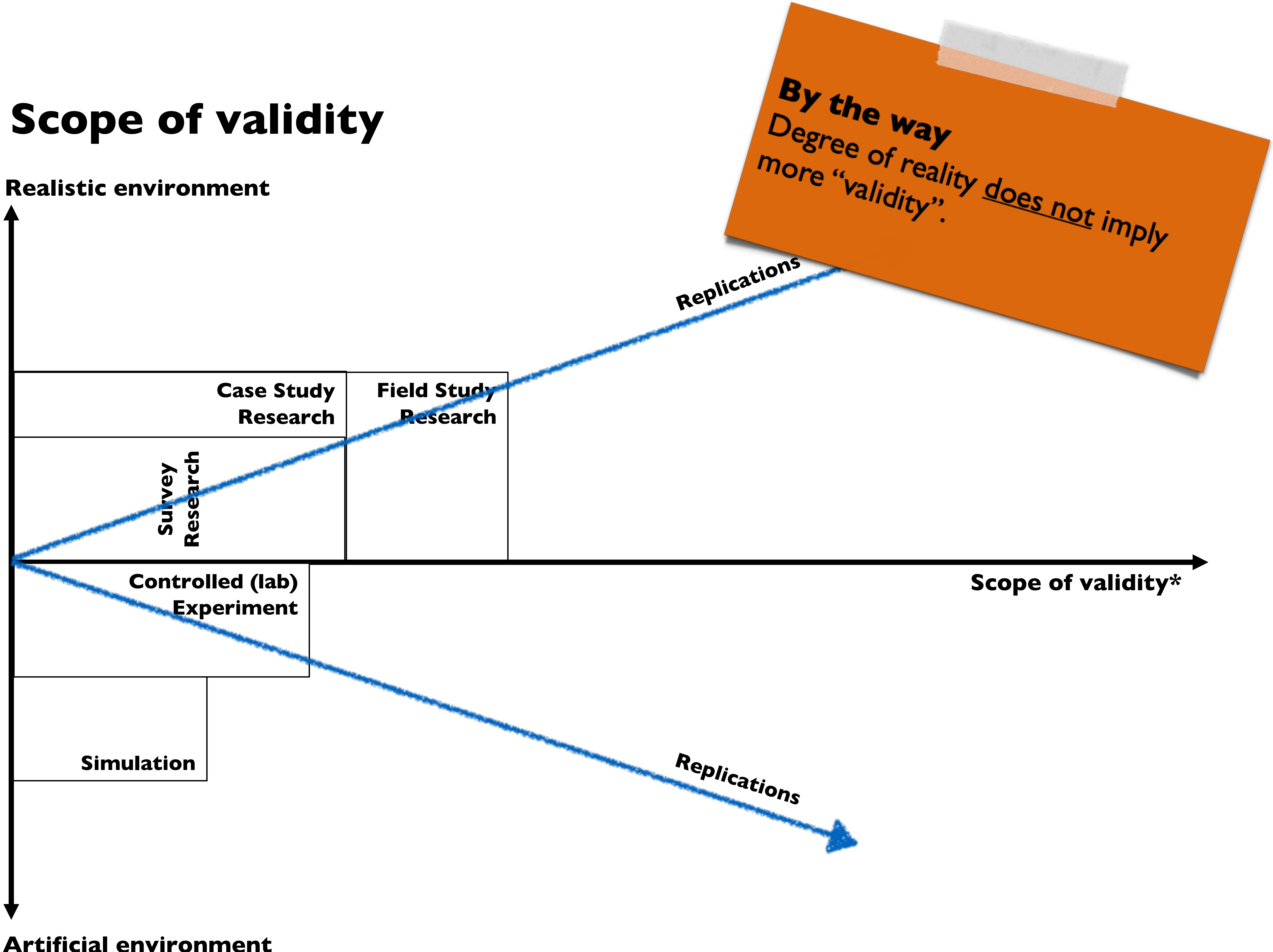
The scope of interest is elaborated in multiple steps

Every study has a specific scope of validity!



Source: Sjøberg, D., Dybå, T., Anda, B., Hannay, J. Building Theories in Software Engineering, 2010.

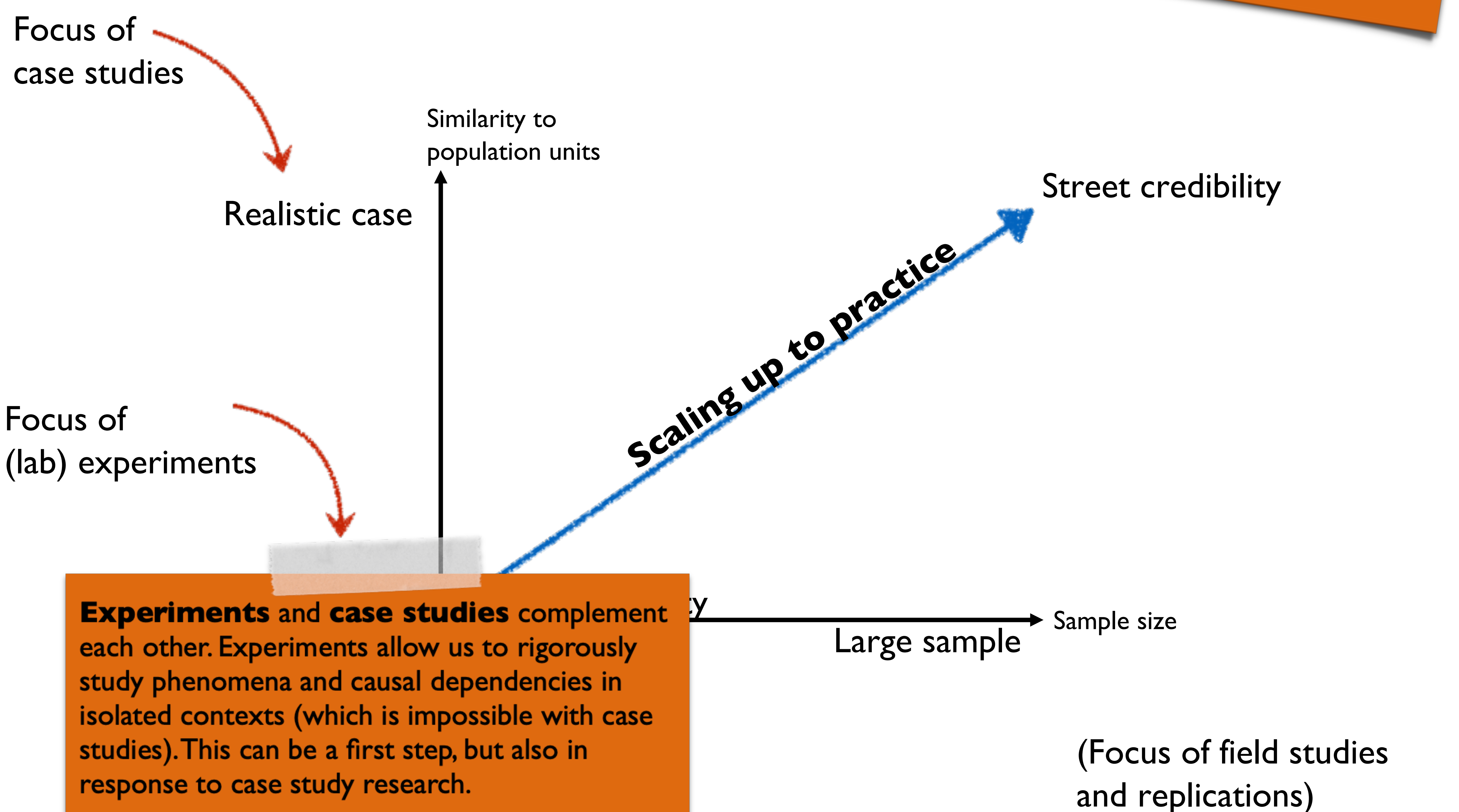
Scope of validity



* Extremely simplified view to orient discussions, please don't sue me.

Case studies and experiments complement each other in scaling up to practice

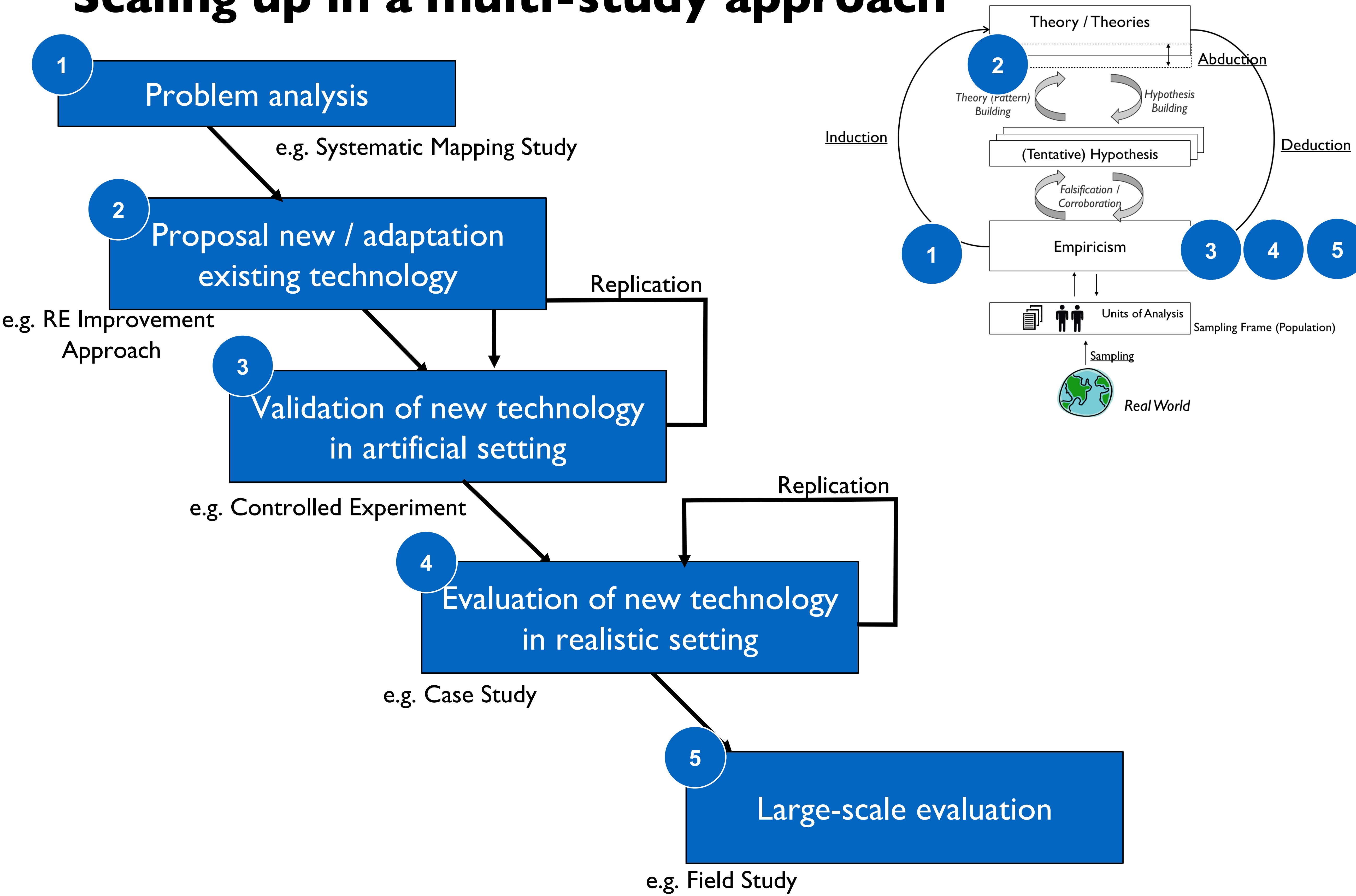
Excursion



Experiments and **case studies** complement each other. Experiments allow us to rigorously study phenomena and causal dependencies in isolated contexts (which is impossible with case studies). This can be a first step, but also in response to case study research.

Based on: Wieringa R. Empirical Research Methods for Technology Validation: Scaling Up to Practice, 2013.

Scaling up in a multi-study approach



Outline

- Science (in a Nutshell)
- Philosophy of Science - a Historical Perspective
- Key Take Aways
- From Philosophy of Science to Empirical Software Engineering
- Empirical Software Engineering Processes
- **Current Challenges in Empirical Software Engineering**

Background: ISERN (Community)

Foto taken at ISERN meeting 2016 in Ciudad Real, Spain

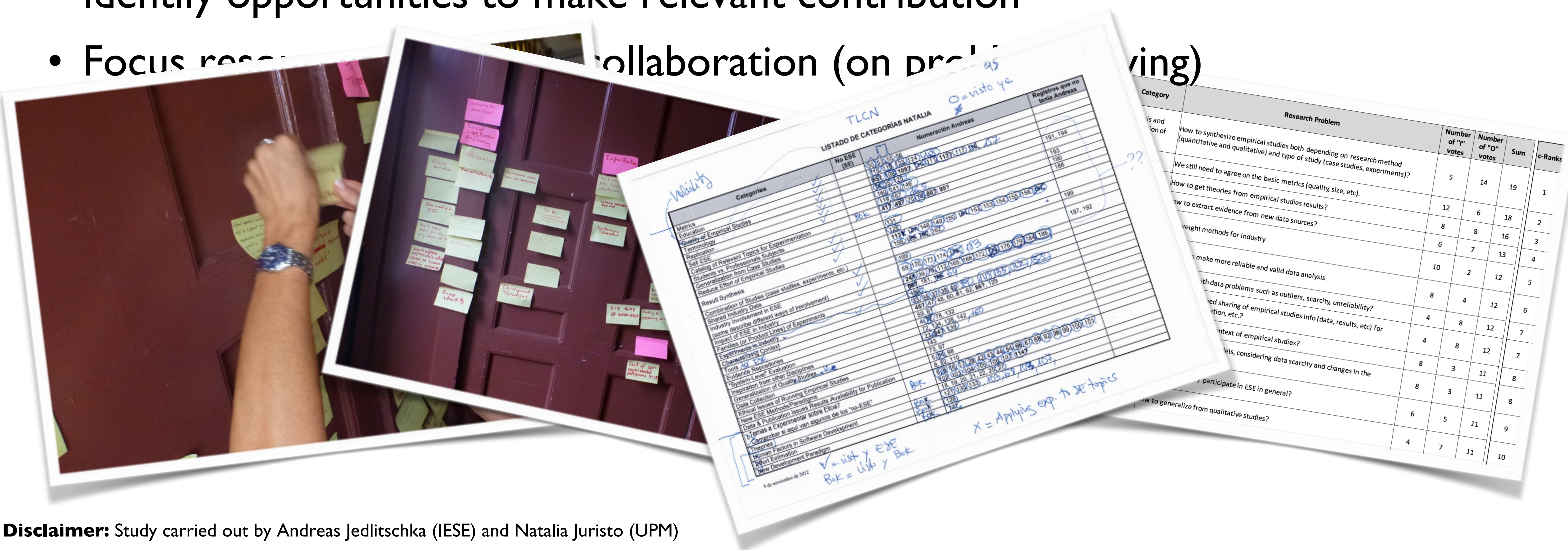
Background (cont.)

Goal

Agree, within the community, on top problems in empirical software engineering

Reason

- Boost a common understanding of what we think is important
- Identify opportunities to make relevant contribution
- Focus reso… collaboration (on probl… ving)



Disclaimer: Study carried out by Andreas Jedlitschka (IESE) and Natalia Juristo (UPM)

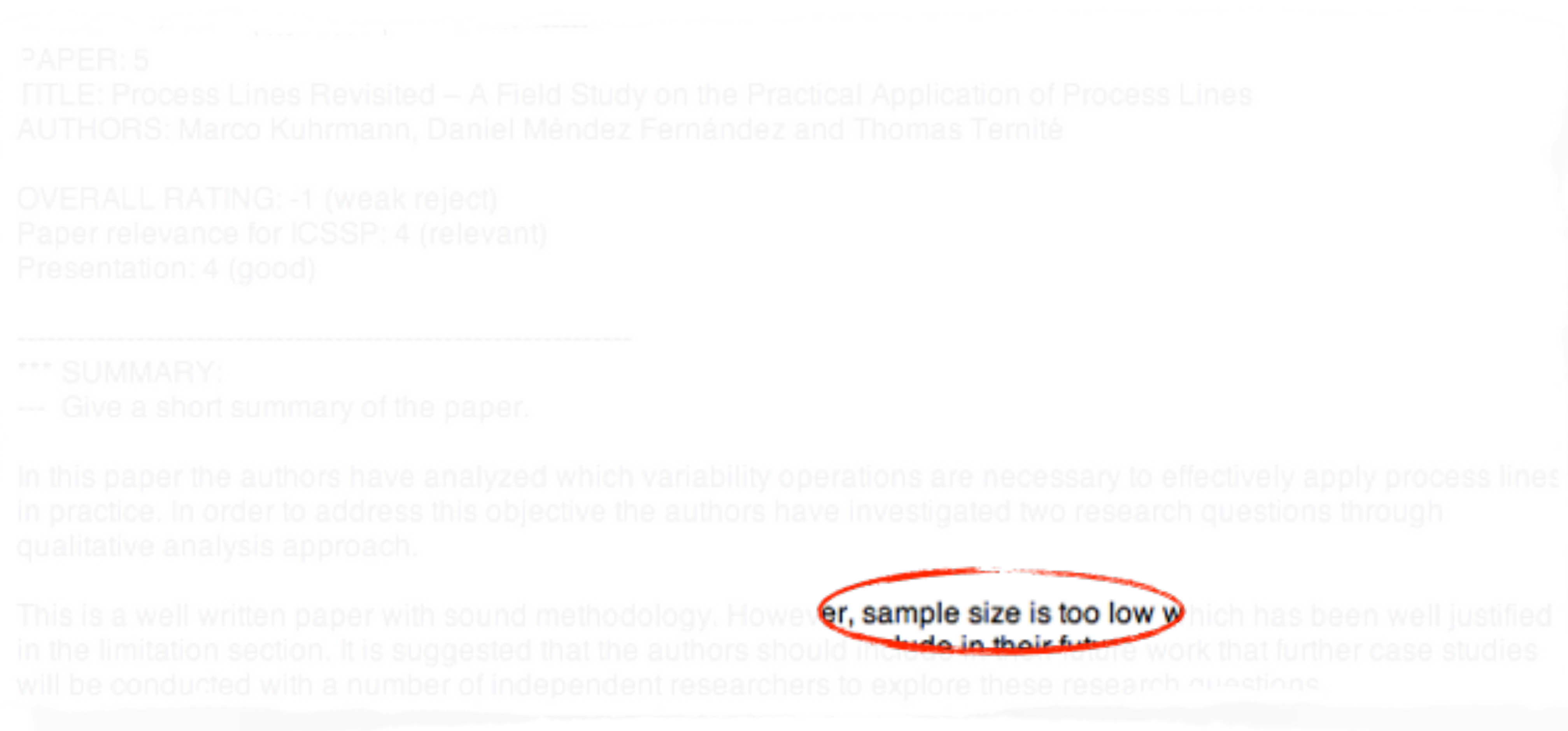
Top challenges as experienced by ISERN members (excerpt)

In total...
... they revealed approx. 150 detailed problems we face in our community (details on demand)

- Collaboration with industry / reaching out
- Contextualisation
- Data collection (incl. automation) and data quality assurance
- Generalisation of results and theory building
- Families of studies and replications
- (Unified) terminology
- Sampling
- Quality (assurance) of empirical studies
- Synthesis and aggregation of results
- (Standardised) reporting

Example!

Sampling: size (doesn't) matter


PAPER: 5
TITLE: Process Lines Revisited – A Field Study on the Practical Application of Process Lines
AUTHORS: Marco Kuhrmann, Daniel Méndez Fernández and Thomas Ternité

OVERALL RATING: -1 (weak reject)
Paper relevance for ICSSP: 4 (relevant)
Presentation: 4 (good)

*** SUMMARY:
--- Give a short summary of the paper.

In this paper the authors have analyzed which variability operations are necessary to effectively apply process lines in practice. In order to address this objective the authors have investigated two research questions through qualitative analysis approach.

This is a well written paper with sound methodology. However, sample size is too low which has been well justified in the limitation section. It is suggested that the authors should include in their future work that further case studies will be conducted with a number of independent researchers to explore these research questions

Too often…

- Sample size seems to be everything

Whereas…

- Appropriateness of sample size depends on research methods employed, and
- the actual population and units of analysis.

Source: Don't ask…

Contextualisation

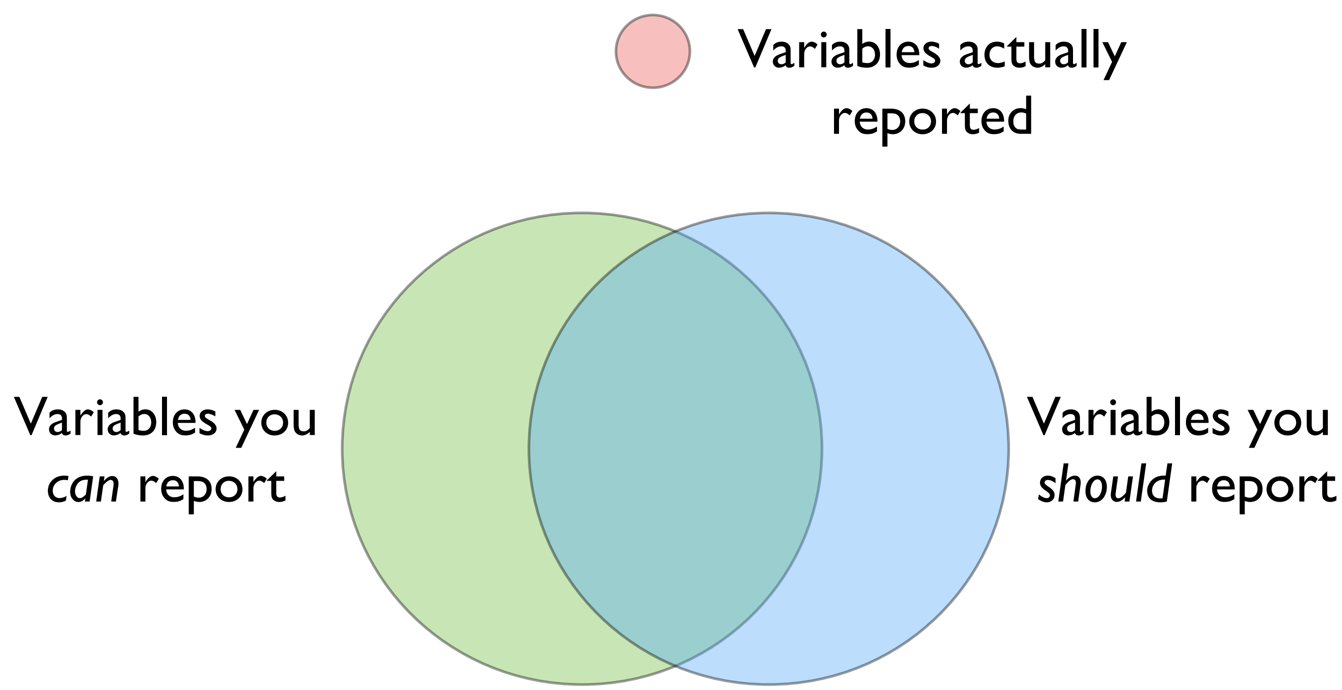
Example!

Too often…

- Researchers do not contextualise properly (or report on it) because
  - it is difficult, or because
  - they are unaware of it.

Whereas…

- The context determines the scope of validity



We have valuable taxonomies with context factors (e.g. in the field of SW process modelling), but they are often unknown or neglected

Example!

Relation to existing evidence

"Close enough. Let's go."

Too often…

- Researchers do often not report on the relation to existing evidence

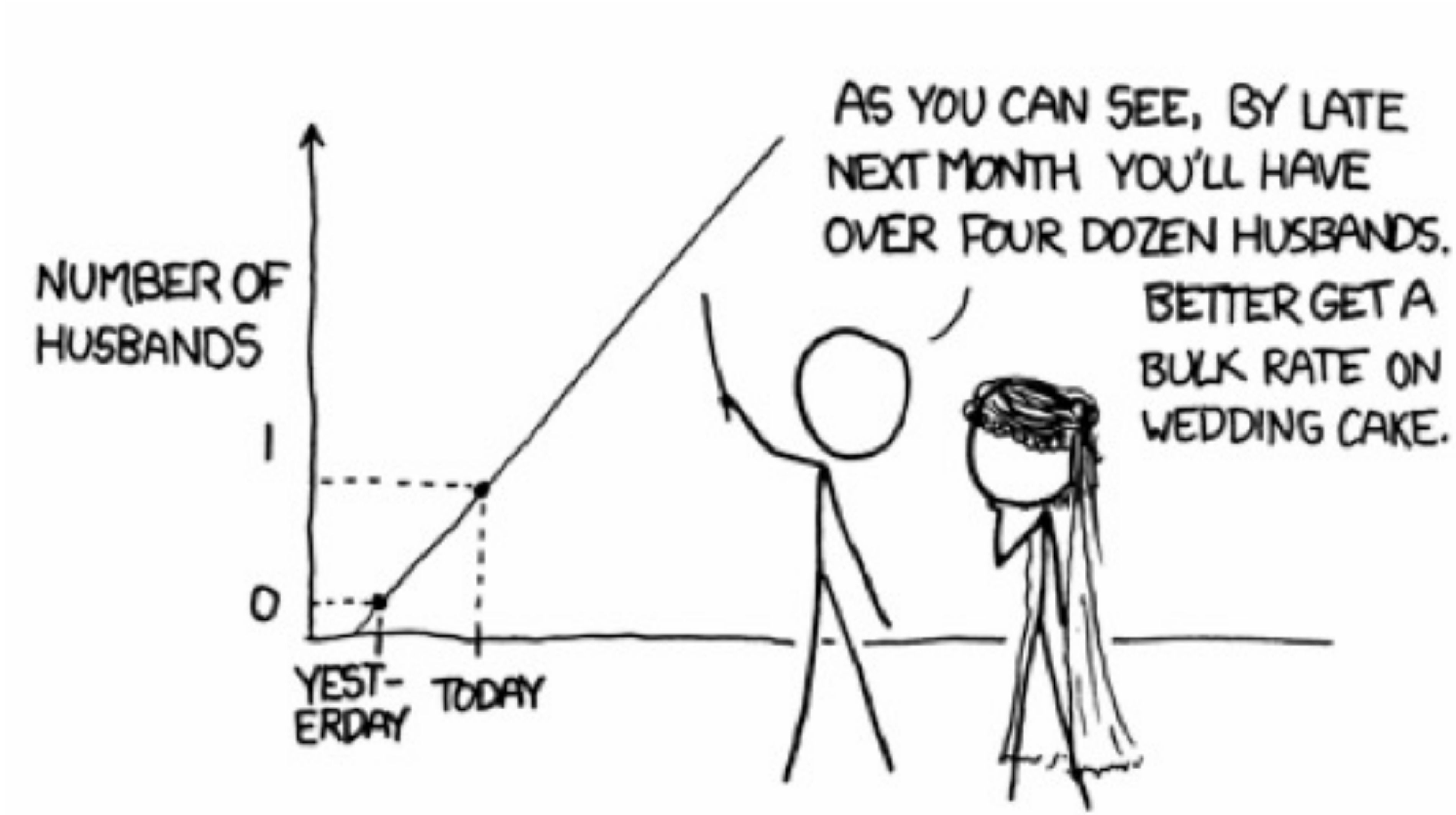
Whereas…

- Scientific progress depends on exactly this

Image source: Huff, D. How to lie with statistics, 1954.

Example!

Conclusion drawing



Too often…

- People mistake knowledge with statistical significance while over-interpreting their results

Whereas…

- Progress is a step-wise procedure (while every step has its own value)

Image source: xkcd

Accurate reporting of results

Example!

Too often…

- Results are not properly reported / disseminated to practice and / or academia
  - lack of details
  - lack of structure

Whereas…

- Proper reporting is the foundation for reliability and reproducibility

  We have standards for reporting study results, but they are often unknown or simply neglected

Image source: [Shameful paper that shall not be named]

(Data) Openness

Example!

Too often…

- Researchers do not share their data, because they might be
  - afraid of doing so
  - not able to do so (NDAs)
  - unaware of the possibilities to do so
  - …

Whereas…

- Openness is the foundation for
  - reliability and trustworthiness
  - reproducibility and replicability

  We have repositories for sharing the data

ANNOUNCEMENT

Reducing our irreproducibility

Over the past year, *Nature* has published a string of articles that highlight failures in the reliability and reproducibility of published research (collected and freely available at go.nature.com/huhbyr). The problems arise in laboratories, but journals such as this one compound them when they fail to exert sufficient scrutiny over the results that they publish, and when they do not publish enough information for other researchers to assess results properly.

From next month, *Nature* and the Nature research journals will introduce editorial measures to address the problem by improving the consistency and quality of reporting in life-sciences articles. To ease the interpretation and improve the reliability of published results we will more systematically ensure that key methodological details are reported, and we will give more space to methods sections. We will examine statistics more closely and encourage authors to be transparent, for example by including their raw data.

Central to this initiative is a checklist intended to prompt authors to disclose technical and statistical information in their submissions, and to encourage referees to consider aspects important for research reproducibility (go.nature.com/oloeip). It was developed after discussions with researchers on the problems that lead to irreproducibility, including workshops organized last year by US National Institutes of Health (NIH) institutes. It also draws on published concerns about reporting standards (or the lack of them) and the collective experience of editors at Nature journals.

The checklist is not exhaustive. It focuses on a few experimental and analytical design elements that are crucial for the interpretation of research results but are often reported incompletely. For example, authors will need to describe methodological parameters that can introduce bias or influence robustness, and provide precise characterization of key reagents that may be subject to biological variability, such as cell lines and antibodies. The checklist also consolidates existing policies about data deposition and presentation.

We will also demand more precise descriptions of statistics, and we will commission statisticians as consultants on certain papers, at the editor's discretion and at the referees' suggestion.

We recognize that there is no single way to conduct an experimental study. Exploratory investigations cannot be done with the same level of statistical rigour as hypothesis-testing studies. Few academic laboratories have the means to perform the level of validation required, for example, to translate a finding from the laboratory to the clinic. However, that should not stand in the way of a full report of how a study was designed, conducted and analysed that will allow reviewers and readers to adequately interpret and build on the results.

To allow authors to describe their experimental design and methods in as much detail as necessary, the participating journals, including *Nature*, will abolish space restrictions on the methods section.

To further increase transparency, we will encourage authors to provide tables of the data behind graphs and figures. This builds on our established data-deposition policy for specific experiments and large data sets. The source data will be made available directly from the figure legend, for easy access. We continue to encourage authors to share detailed methods and reagent descriptions by depositing protocols in Protocol Exchange (www.nature.com/protocolexchange), an open resource linked from the primary paper.

Renewed attention to reporting and transparency is a small step. Much bigger underlying issues contribute to the problem, and are beyond the reach of journals alone. Too few biologists receive adequate training in statistics and other quantitative aspects of their subject. Mentoring of young scientists on matters of rigour and transparency is inconsistent at best. In academia, the ever increasing pressures to publish and chase funds provide little incentive to pursue studies and publish results that contradict or confirm previous papers. Those who document the validity or irreproducibility of a published piece of work seldom get a welcome from journals and funders, even as money and effort are wasted on false assumptions.

Tackling these issues is a long-term endeavour that will require the commitment of funders, institutions, researchers and publishers. It is encouraging that NIH institutes have led community discussions on this topic and are considering their own recommendations. We urge others to take note of these and of our initiatives, and do whatever they can to improve research reproducibility. ■

figshare
credit for **all** your research

promisedata
Data for reusable SE experiments

Source: http://www.nature.com/news/announcement-reducing-our-irreproducibility-1.12852

How to deal with all these problems?

- Education on scientific working including
  - research methods and practices - The What
  - their setting in a bigger picture - The Why
- Rely on standards and contribute to standards
- Take your message out (evangelise)

Thank you!

(and enjoy the rest of the course :-)

Daniel Méndez
Daniel.Mendez@tum.de
@mendezfe