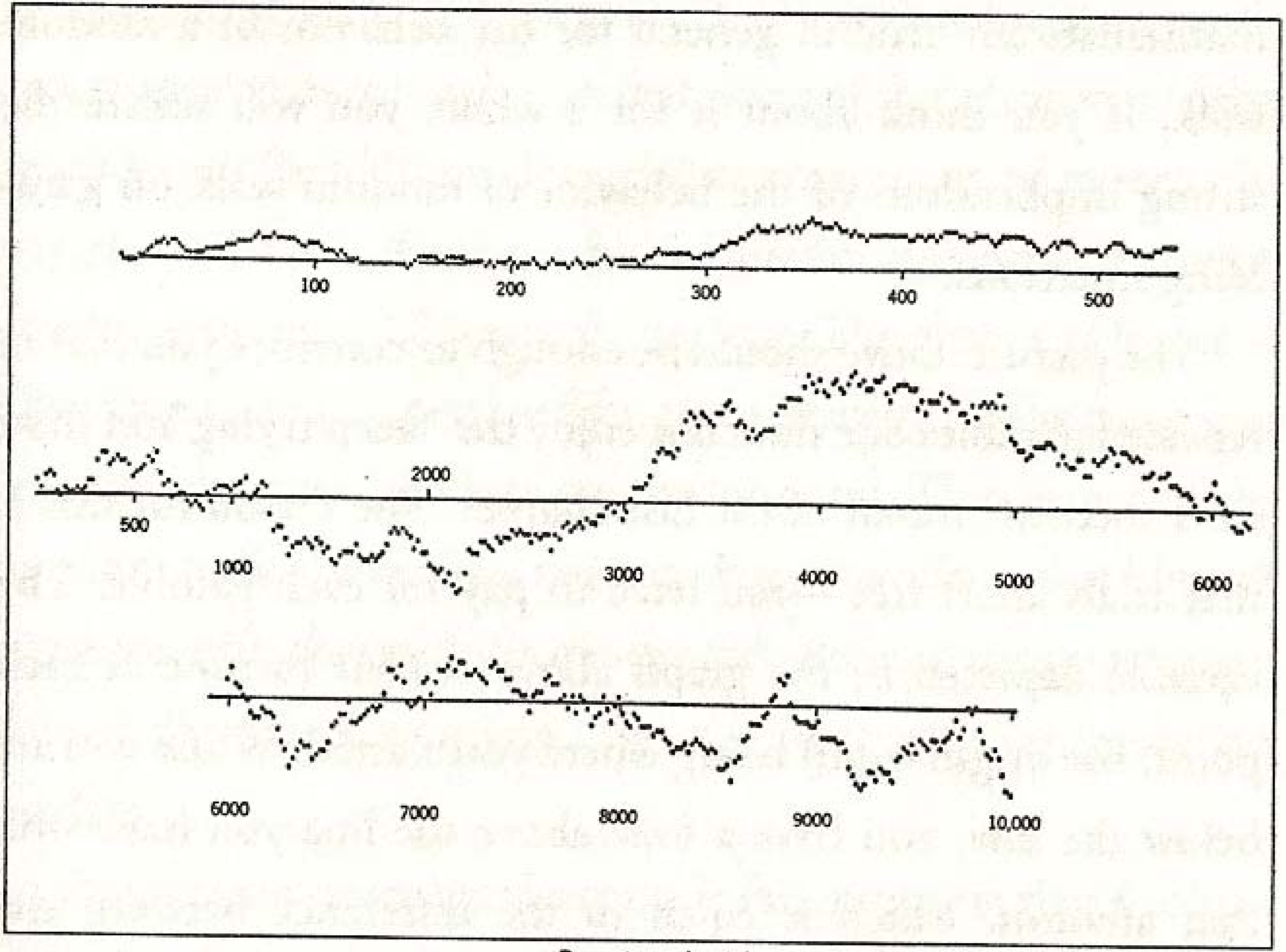
The End of Theory?
On the role of theories in Information and Knowledge Systems research

Prof. Dr. Roel Wieringa
University of Twente,
The Netherlands
www.cs.utwente.nl/~roelw

© R.J. Wieringa

- 10:30 – 12:00 design cycle, theories, research setup
- lunch
- 12:30 – 14:00 description, abduction, analogy
- tea
- 14:15 – 15:45 statistical inference
- tea
- 16:00 – 16:30 checklist

- Big data allows computation of predictive theories.
  - Such a theory has more credibility if it is based not only on statistics, but also on substantial explanation in terms of underlying mechanisms.



Reprinted with permission of John Wiley & Sons, Inc.

- Mindless statistics, fancy philosophy and impressive machines
  - Many SE papers use mindless statistics to impress their peers and reject a silly null hypothesis without further explanation.
  - Many IS papers use fancy philosophy to impress their peers and present trivial insight as grand theory.
  - Many AI papers display technical prowess in conference papers,
- All of these aim to impress their peers, just as Harley Davidson fanatics show each other how they pimped up their motorcycle at yearly gatherings.

THREE-WHEELER
WATERCOOLED
LAŠKO
LJ VR 1

LAŠKO
PIVOVARNA LAŠKO
SCREW IT, LET'S RIDE
Cockta

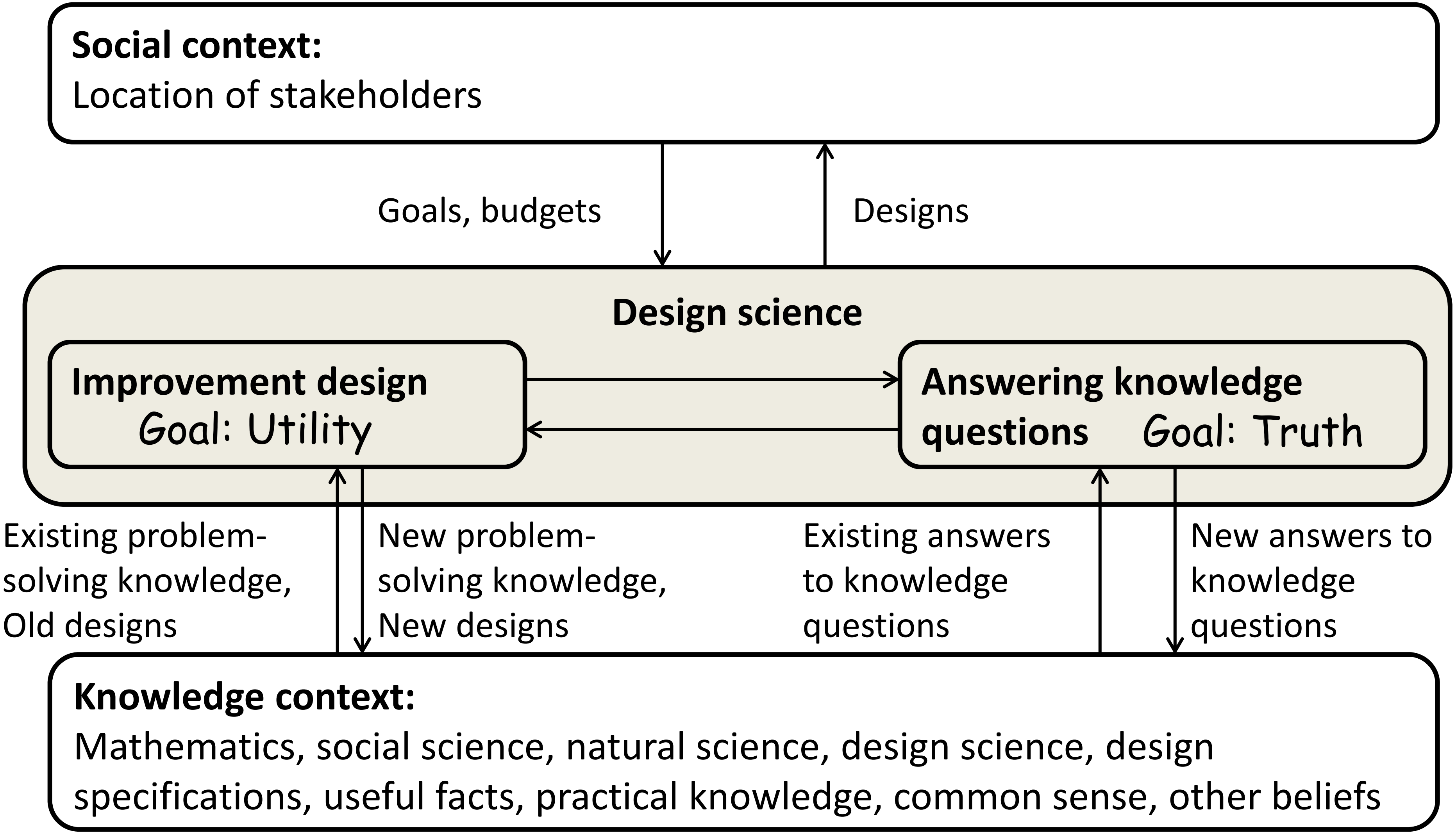
Outline

- **Introduction**
  - The design cycle
  - Theories
  - The research setup
- **Scientific inference**
  - Description
  - Explanation
  - Generalization
  - Prediction
- **Empirical research**
  - Checklist for research, reading papers, and writing papers
  - Example research methods

Outline

- **Introduction**
  - The design cycle
  - Theories
  - The research setup
- **Scientific inference**
  - Description
  - Explanation
  - Generalization
  - Prediction
- **Empirical research**
  - Checklist
  - Example research methods

Framework for design science



Social context:
Location of stakeholders

Goals, budgets

Designs

Design science

Improvement design
Goal: Utility

Answering knowledge questions Goal: Truth

Existing problem-solving knowledge, Old designs

New problem-solving knowledge, New designs

Existing answers to knowledge questions

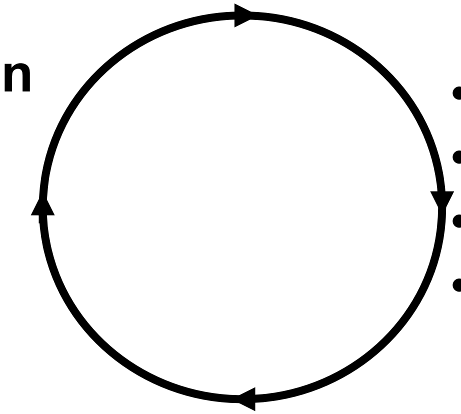
New answers to knowledge questions

Knowledge context:
Mathematics, social science, natural science, design science, design specifications, useful facts, practical knowledge, common sense, other beliefs

Engineering cycle

This is a checklist. See appendix A in the book & on my web site

! = Action
? = Knowledge question



Treatment implementation

Implementation evaluation = Problem investigation

- Stakeholders? Goals?
- Conceptual problem framework?
- Phenomena? Causes, mechanisms, reasons?
- Effects? Positive/negative goal contribution?

Treatment validation

- Context & Artifact → Effects?
- Effects satisfy Requirements?
- Trade-offs for different artifacts?
- Sensitivity for different Contexts?

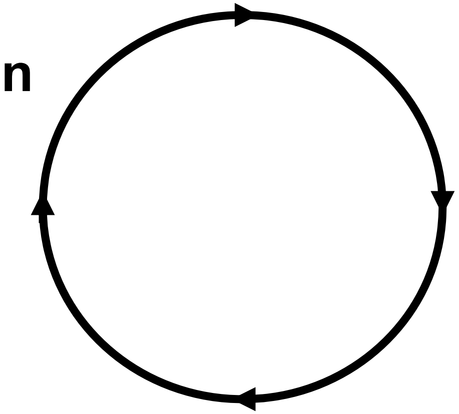
Treatment design

- Specify requirements!
- Requirements contribute to goals?
- Available treatments?
- Design new ones!

Engineering cycle **in the laboratory**

Implementation evaluation = Problem investigation

- *Researchers want to explore a design*
- *Conceptual problem framework to specify the design: Defined in research papers*
- *Phenomena: Performance data, explanations of these*



Treatment implementation

- *Build a prototype and a test environment; run it*

Treatment validation

- *Predict effects in a context*
- *Compare with requirements*
- *Compare with other designs*
- *Check assumptions about context*

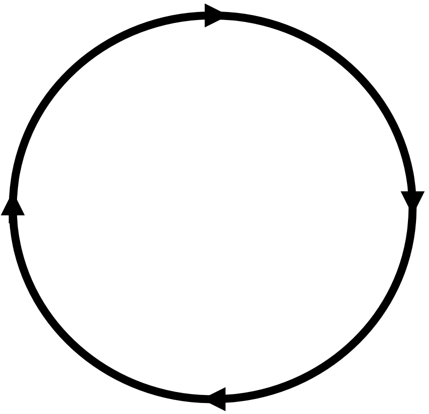
Treatment design

- *Specify required performance*
- *Motivate in terms of design goals*
- *Consider existing designs*
- *Design a new one*

Engineering cycle **in the real world**

Treatment implementation

- *Transfer to market*



Implementation evaluation = Problem investigation

- *Real world stakeholders want to achieve goals*
- *They conceptualize the world in some way*
- *Problems are experienced, and (mis)understood*
- *These problems have undesirable effects*

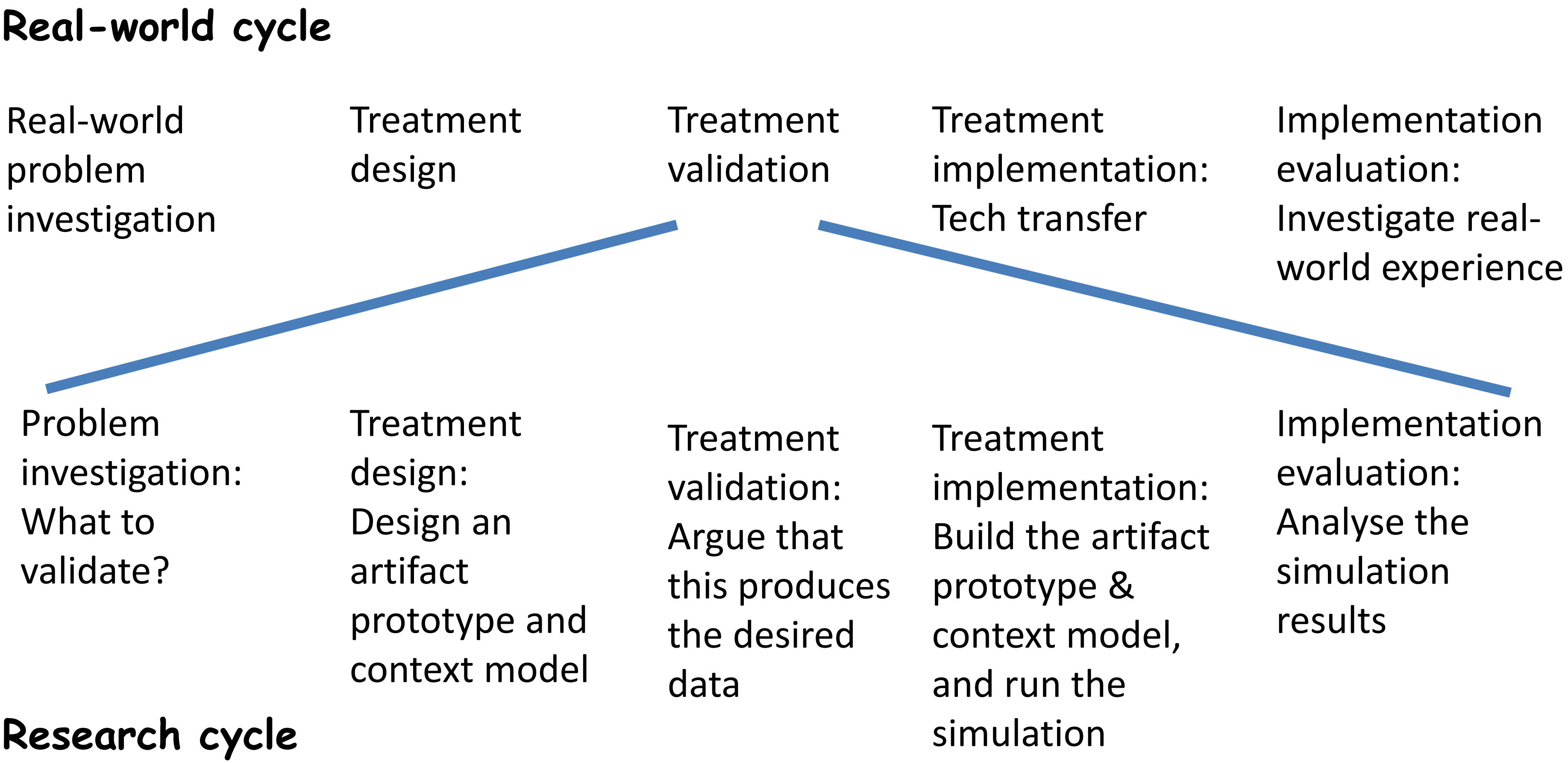
Previous slide

Treatment validation

- *Predict effects in a context by lab experimentation*
- *Compare with requirements*
- *Compare with other designs*
- *Check assumptions about context*

Treatment design

- *Specify required performance*
- *Motivate in terms of stakeholder goals*
- *Consider existing solutions*
- *Design a new one*



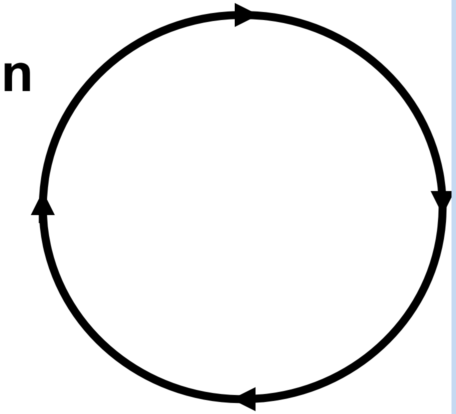
Real-world cycle

Real-world problem investigation	Treatment design	Treatment validation	Treatment implementation: Tech transfer	Implementation evaluation: Investigate real-world experience
Problem investigation: What to validate?	Treatment design: Design an artifact prototype and context model	Treatment validation: Argue that this produces the desired data	Treatment implementation: Build the artifact prototype & context model, and run the simulation	Implementation evaluation: Analyse the simulation results

Research cycle

- Henceforth, “engineering cycle” means “real-world engineering cycle”.
- The research cycle will emerge as empirical research cycle.

Engineering cycle
in the real world



Treatment implementation

- *Transfer to market*

Implementation evaluation = Problem investigation

Development of **problem theories** about stakeholders and their problems, or of **design theories** about artifacts and their real-world behavior, based on real-world observations

Treatment design

Design some artifact

Treatment validation

Development of **design theories** about artifacts and their real-world behavior, based on simulations

K. Peffers, T. Tuunanen, M.A. Rothenberger, S. Chatterjee, A design science research methodology for information systems research. J. Manag. Inf. Syst. **24**(3), 45–77 (2007–2008)

Peffers et al	**Design cycle**
Problem identification and motivation	Problem investigation
Objectives of a solution	Treatment design: specify requirements
Design . . .	Treatment design: the rest
. . . and development	Validation: instrument development. Develop prototype and model of context
Demonstration	Validation: effects, trade-offs, sensitivity?
Evaluation	Validation: do effects satisfy requirements?
Communication	

- **Problem theories** are about stakeholders and their goals and problems
  - Theories from psychology, sociology, economics, management science
  - *Theory of cognitive dissonance*
    - *Inconsistent cognitions are uncomfortable. People change this by*
    - *(1) changing their behavior,*
    - *(2) promising to change their behavior,*
    - *(3) changing the norms applicable to behavior,*
    - *(4) denying the laws of nature.*
  - *Balance theorem in social networks*
    - *A complete network with only +++ and +-- triangles partitions into two giant subnetworks who internal like each other and externally hate each other.*
  - *Transaction cost theory*
    - *Firms exist to reduce transaction cost*

- **Design theories** are about artifacts in context
  - *RE in agile projects for SME's is done by developers … because the SME will not make resources available for SW development.*
  - *SW project effort estimations in our bank are too low …. because not all requirements are known.*
  - *Our new modeling method is usable and useful for domain experts … because it does not require learning and allows them to express their knowledge.*
  - *Our new route planning algorithm produces less delays on airports than fixed planning …. because it responds to traffic jams and the airport road network has only few starting points and destinations.*
- Observations
  - Design theory are local: about a particular artifact in a particular context
  - Relevance of design theories is context- and technology-dependent
  - Prototypes built to test a design theory will be lost

Problem theories

		Goal of theory-building	
		Problem understanding	Designing
Subject of the theory	Stakeholders, goals, problems	To understand a problem	To justify an intervention in a problem
	Artifact in Context	To evaluate an artifact in a context	To validate the design an artifact for a context

Design theories

Some methodologists take the concept of problem theory wider: They talk about natural science theories

Outline

- **Introduction**
  - The design cycle
  - Theories
  - The research setup
- **Scientific inference**
  - Description
  - Explanation
  - Generalization
  - Prediction
- **Empirical research**
  - Checklist
  - Example research methods

What is a theory?

- A **theory** is a belief that there is a pattern in phenomena.
  - Idealizations: *“Merging two faculties reduces cost.” “This works in theory, but not in practice.”*
  - Speculations: *“Elvis lives.” “Jemenites are all terrorists.” “9/11 was executed by the CIA’”*
  - Opinions: *“The Dutch lost the soccer competition because the players are prima donna’s that do not play like a team.’’*
  - Wishful thinking: *``My technique works better than the others.’’*
  - Scientific theories*: Theory of electromagnetism*

- Theories may be general or particular
  - They may state that there is a pattern
  - They may indicate that a phenomenon is an instance of a pattern

What is a scientific theory?

- A **theory** is a belief that there is a pattern in phenomena.
- A **scientific** theory is a belief that there is a pattern in phenomena, that has survived
  - Tests against experience:
    - Observation, measurement
    - Possibly: experiment, simulation, trials
  - Criticism by critical peers:
    - Anonymous peer review
    - Publication
    - Replication

Non-examples

- *Religious beliefs*
- *Political ideology*
- *Marketing messages*
- *Most social network discussions*

Examples

- *Theory of electromagnetism*
- *Technology acceptance model*

What is a scientific design theory?

- A **theory** is a belief that there is a pattern in phenomena.
- A **scientific** theory is a belief that there is a pattern in phenomena, that has survived
  - Tests against experience,
  - Criticism by critical peers.
- A **scientific design theory** is a belief that there is a pattern in the interaction between an artifact and its context, that has survived tests against experience and criticism by critical peers.

Examples:

- *Theory of the UML in software engineering projects*
- *Theory about accuracy and speed of DOA algorithms in a context of plane waves and white noise*
- *Theory about delays in routes planned by MARP on airports*

S. Gregor, The nature of theory in information systems. MIS Q. **30**(3), 611–642 (2006)

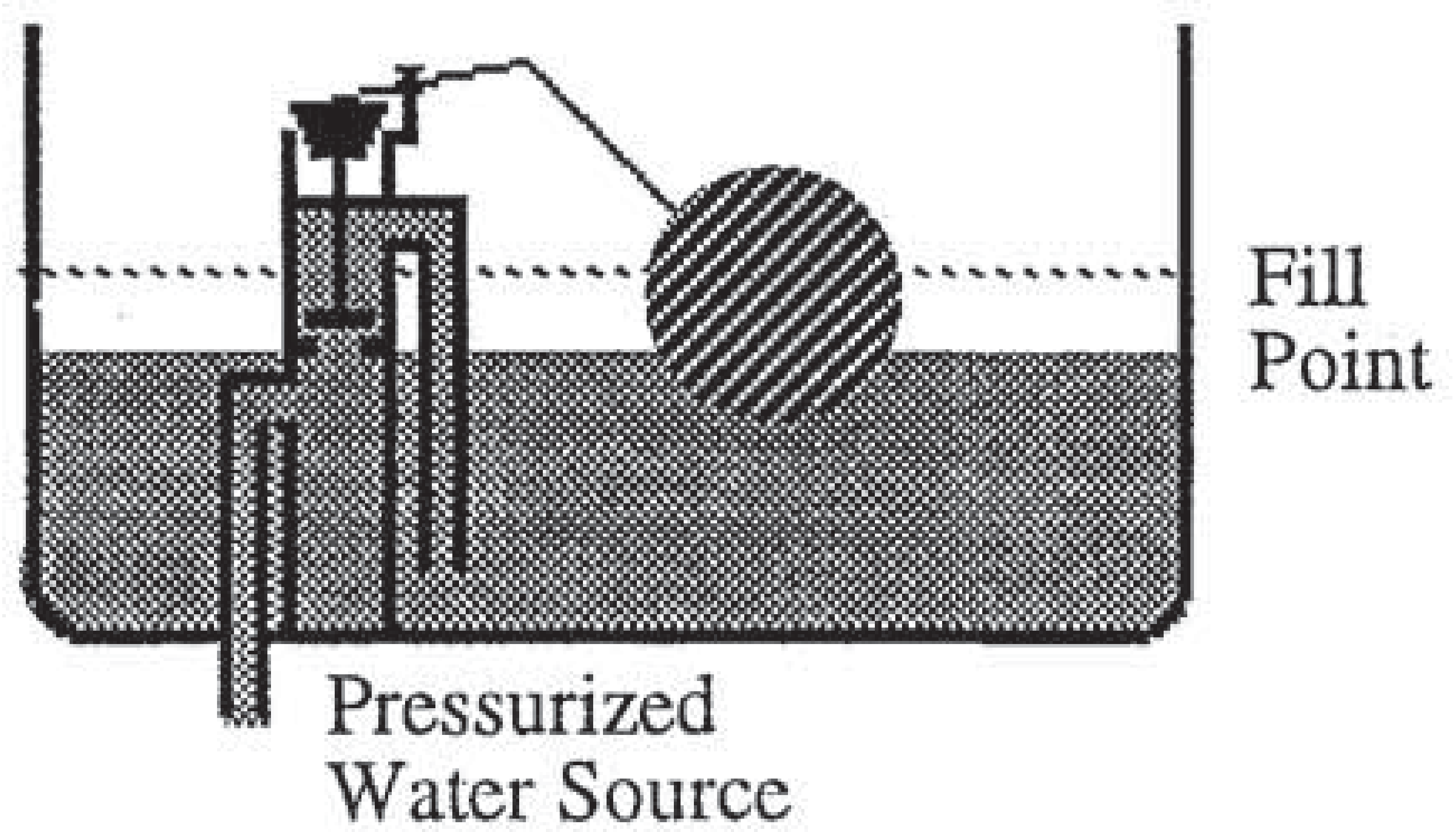
Gregor	**This course**
Means of representation	Not part of a theory in this book. One and the same theory can be represented in many different ways without changing the theory
Constructs	Conceptual framework
Statements of relationship	Generalizations
Scope	Scope
Causal explanation	This is one kind of explanation, next to architectural and rational explanations. Theories may be descriptive too.
Testable propositions (hypotheses)	Theories must be empirically testable, but testable propositions derived from the theory are not part of the theory
Prescriptive statements	Scientific theories do not prescribe anything

Theory

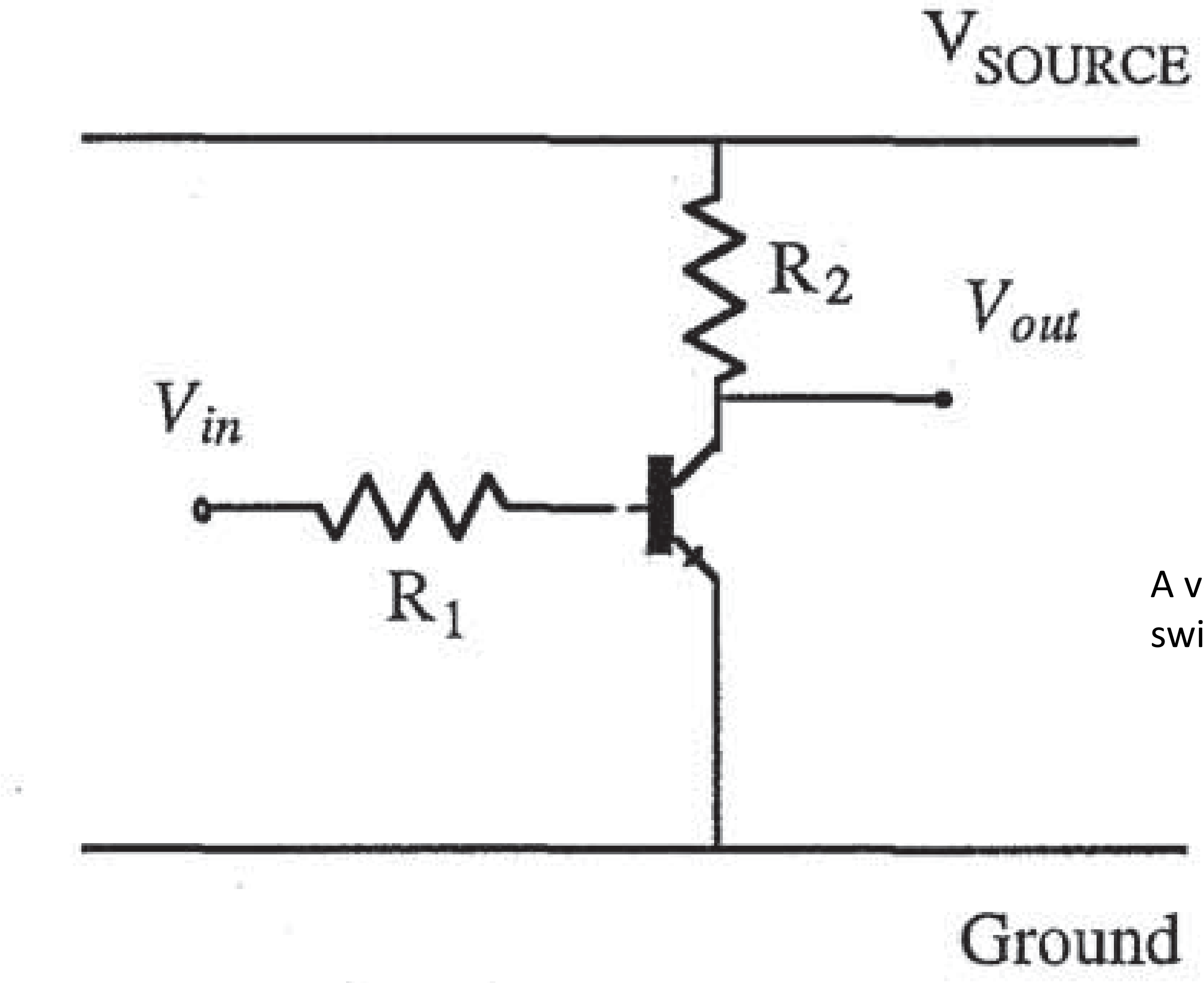
S. Gregor, D. Jones, The anatomy of a design theory. J. AIS **8**(5), 312–335 (2007)

Gregor & Jones	**This course**
Constructs	Conceptual framework
Testable propositions	Generalizations
Scope	Scope
Justificatory knowledge	Prior knowledge
Purpose	Artifact requirements, stakeholder goals
Principles of form and function	Design choices
Artifact mutability	Artifact variants (trade-offs)
Principles of implementation	Could be part of implementation theory
Expository instantiation	Validation model

Not part of a design theory (but part of its engineering cycle context)

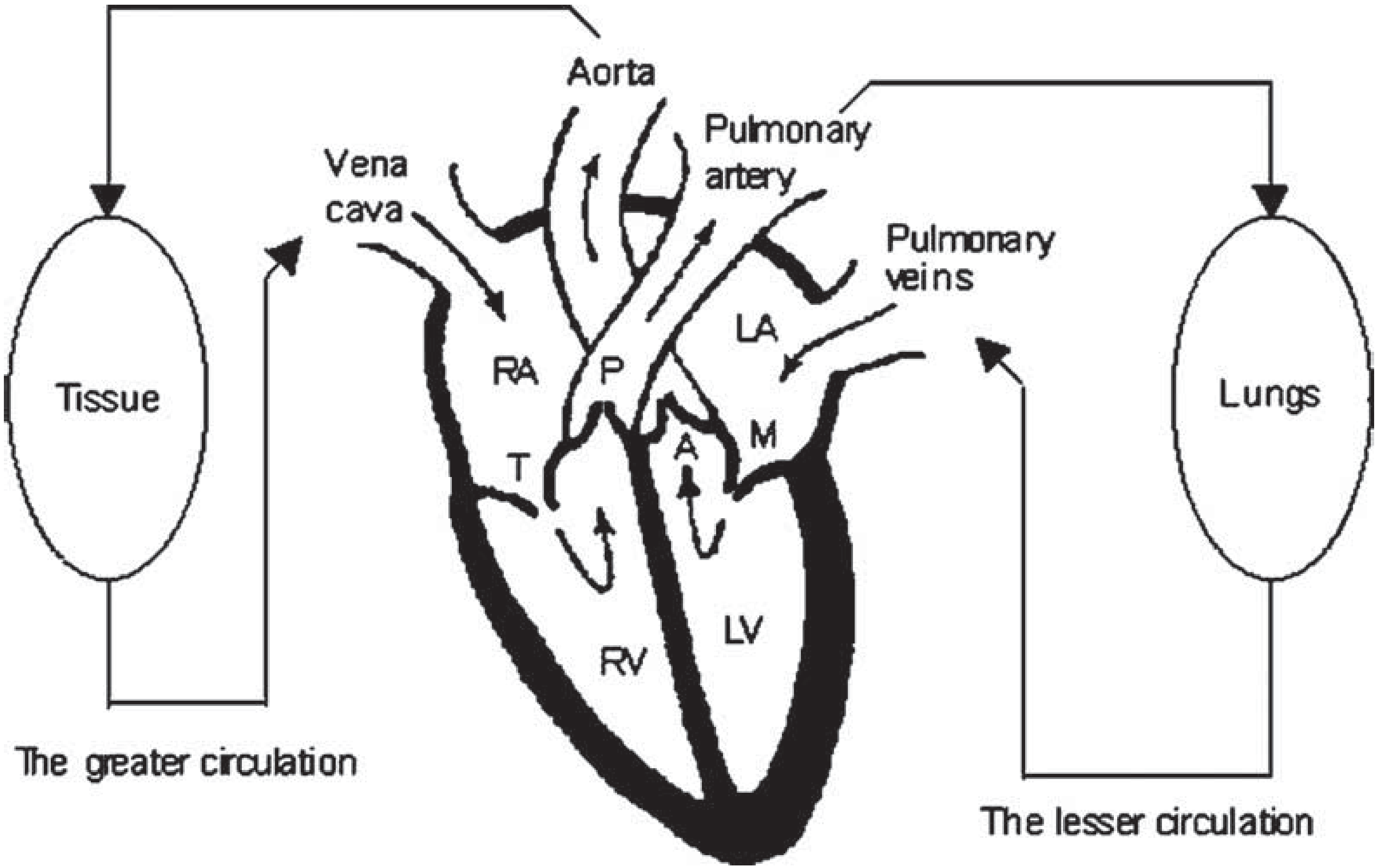


- Conceptual model of an artifact architecture.
- Together with a narrative of the mechanism, this diagram is a design theory of an artifact.

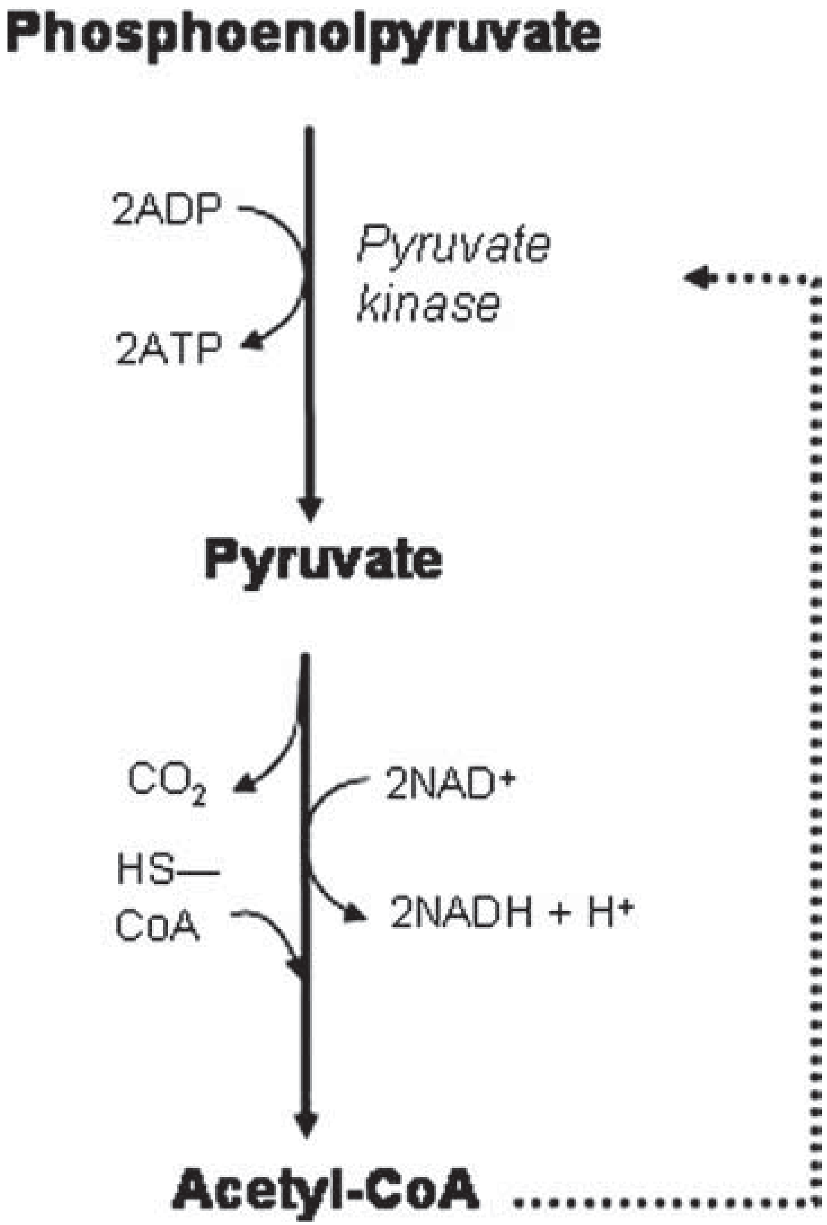


A voltage switch

- Conceptual model of a artifact mechanism.
- Together with a narrative of the mechanism, this diagram is a design theory of an artifact.



- Conceptual model of a natural architecture.
- Together with a narrative of the mechanism, this model is a theory of a natural process.



- Feedback loop in the linkage between two metabolic systems
- Conceptual model of a natural architecture (components and interactions).
- Together with a narrative of the mechanism, this is a theory of a natural process.

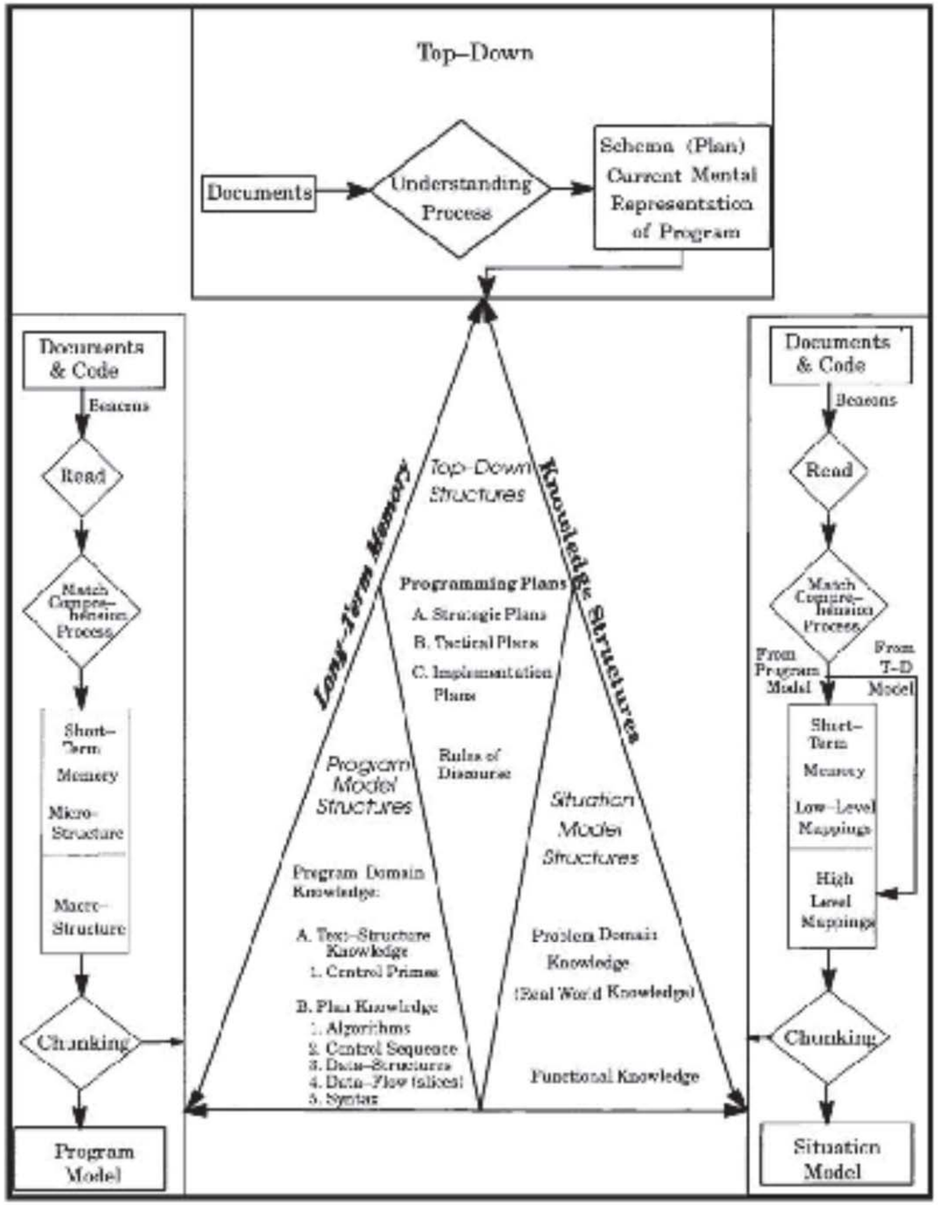


Figure 1. Diagrammatic representation of the Integrated Comprehension model from von Mayrhauser and Vans (1995a)

- Cognitive mechanism of program comprehension
- Conceptual model of a natural architecture
- Together with a narrative of the mechanism, <u>this is a theory </u>of a natural process.

J. Softw. Maint., 9, 299–327 (1997)

© 1997 John Wiley & Sons, Ltd.

The structure of scientific theories

1. **Conceptual framework**
   - Definitions of concepts.
2. **Generalizations**
   - Express (in the form of text, formulas, diagrams) beliefs about patterns in phenomena in a population:
     - Descriptions of a pattern
     - Explanations of a pattern

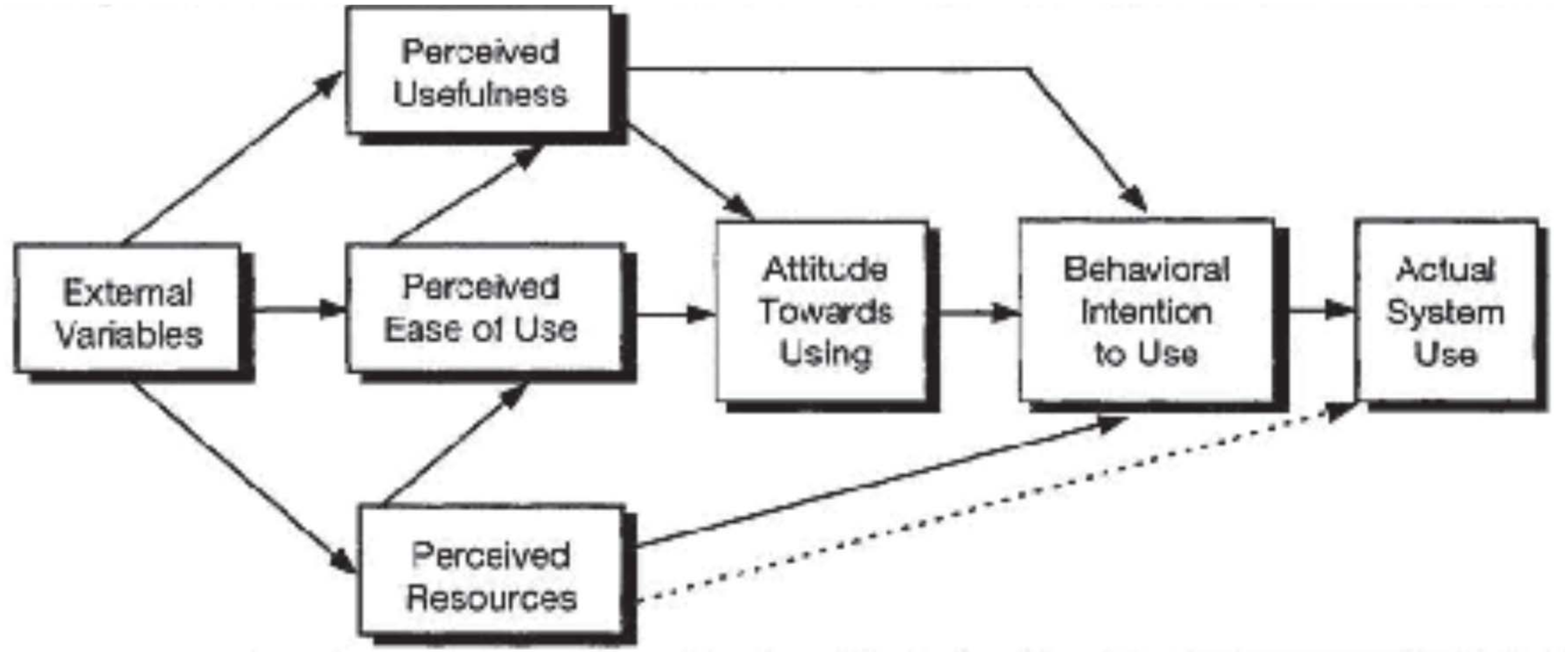
- If generalizations are mathematical, there is an inaccurate match between exact generalizations and inexact real world phenomena

Theory of electromagnetism

- *Conceptual framework (concepts defined to describe and explain the relevant phenomena):*
  - *Electric current, electric charge, potential difference, electric resistance, electric power, capacitance, electric field, magnetic field, magnetic flux density, inductance, …, … and their units.*
- *Generalizations*
  - *Electric charges attract or repel one another with a force inversely proportional to the square of their distance.*
  - *An electric current inside a wire creates a corresponding circular magnetic field outside the wire.*
  - *…*
- Conceptual framework to make architectural models of a class of artificial or natural systems
- Generalizations about mechanisms in those systems
- Use of calculus to quantify propositions

Technology Acceptance Model

- *Conceptual framework*
  - *Definitions of perceived usefulness, perceived ease of use, perceived resources, attitude towards using, behavior intention to use, actual system use*
- *Generalization*



- Conceptual framework with definitions of variables
- Statement of influence relations among these variables

The Balance Theorem in social networks

- *Conceptual framework*
  - *Definition of concepts of graph, link, friend/enemy, complete graph (each pair of nodes connected), balanced graph (no --- or ++- triangles)*
- *Mathematical theorem:*
  - *If a labeled complete graph is balanced, then*
    - *either all pairs of nodes are friends,*
    - *or else the nodes can be divided into two groups, X and Y, such that every pair of nodes in X like each other, every pair of nodes in Y like each other, and everyone in X is the enemy of everyone in Y .*
- Conceptual framework defines a mathematical structure
- Proposition proved in that structure.
- Empirical fact: In the real world, large call networks almost satisfy the assumptions and in fact are almost balanced

Theory of cognitive dissonance

- *Conceptual framework*
  - *Beliefs, intentions, values, facts, observations, conflict between facts and observations*
  - *Capabilities of people: They can …*
    - *Change their behavior*
    - *change their values*
    - *change their intention*
    - *deny observation*
    - *deny fact*
- *Generalization:*
  - *People seek consistency among their cognitions. They resolve this by changing their behavior, changing their values, making promises, ignoring observations, or denying facts.*
- Conceptual framework defines some variables
- Generalization describes a mechanism that often occurs

Theory of the UML

- *Concepts: UML concepts, definitions of software project, of software error, project effort, definition of concept of domain, understandability*
- *Descriptive generalization: (UML) X (SE project) → (Less errors, less effort than similar non-UML projects)*
- *Explanatory generalizations:*
  - *UML models resemble the domain more than other kinds of models;*
  - *They are easier to understand for software engineers;*
  - *So they they make less errors and there is less rework (implying less effort).*

- When you design a new artifact, you (should) have a theory about it
  - What effects it will have
  - Why these happen

- The artifact usually disappears, the theory should stay

Functions of theories

- Functions of a conceptual framework
  - **Framing** a problem or artifact (select words to describe them)
  - **Describe** a problem or
  - **Specify** an artifact
  - **Analyze** a problem or artifact
  - **Generalize** about the problem or artifact
- Functions of a generalization
  - Descriptive generalizations allow us to **predict**
  - Explanatory generalizations allow us to **understand**

Usability of design theories

- When is a design theory

  Context assumptions X Artifact design → Effects

  usable by a practitioner?

  1. The theory must be predictive.
  2. The practitioner is capable to recognize Context Assumptions
  3. and to acquire/build and use the Artifact,
  4. effects will indeed occur when used, and
  5. effects will contribute to stakeholder goals

- Practitioner has to asses the risk that each of these fails

Ucare

- *(Assumptions about elderly and their context) X (Ucare specification) → (Cheaper and better home care)*
- *Usable by a practitioner?*
  1. *It is a predictive theory*
  2. *He/she is capable to recognize Context Assumptions*
  3. *And to acquire/build and use the Artifact,*
  4. *Effects will indeed occur when used, and*
  5. *Effects will contribute to stakeholder goals*

- *What are the risks?*

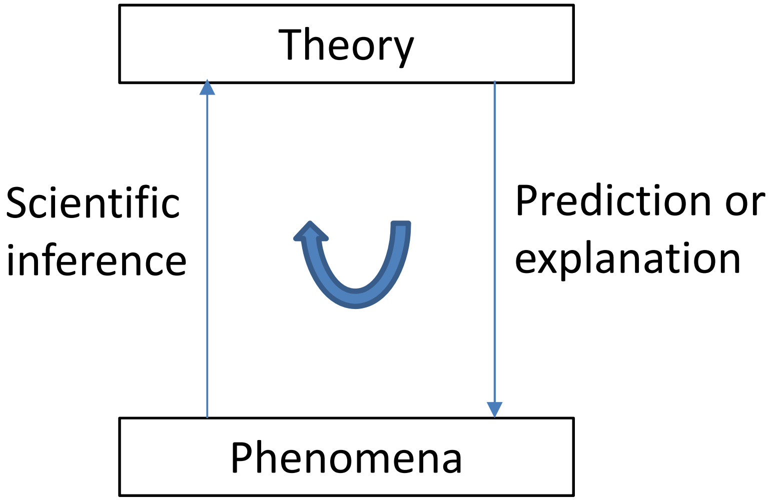
S. Gregor, The nature of theory in information systems. MIS Q. **30**(3), 611–642 (2006)

Theory Type	**Distinguishing Attributes**
I. Analysis	Says what is. The theory does not extend beyond analysis and description. No causal relationships among phenomena are specified and no predictions are made.
II. Explanation	Says what is, how, why, when, and where. The theory provides explanations but does not aim to predict with any precision. There are no testable propositions.
III. Prediction Says what is and what will be.	The theory provides predictions and has testable propositions but does not have well-developed justificatory causal explanations.
IV. Explanation and prediction (EP)	Says what is, how, why, when, where, and what will be.
V. Design and action	Says how to do something. The theory gives explicit prescriptions for constructing an artifact.

S. Gregor, The nature of theory in information systems. MIS Q. **30**(3), 611–642 (2006) **compared with my approach**

Theory Type	Distinguishing Attributes
I. Analysis **Descriptive theory**	**Says what is.** The theory does not extend beyond analysis and description. No causal relationships among phenomena are specified and no predictions are made.
II. Explanation **Explanatory theory**	**Says** what is, how, **why,** when, and where. The theory provides explanations but does not aim to predict with any precision. There are no testable propositions.
III. Prediction Says what is and what will be. **Predictive theory**	**The theory provides predictions** and has testable propositions but does not have well-developed justificatory causal explanations.
IV. Explanation and prediction (EP)	Says what is, how, why, when, where, and what will be.
V. Design and action	Says how to do something. The theory gives explicit prescriptions for constructing an artifact.
Usable theory	**Says how to achieve an effect**

Development and maintenance of theories



- Theories are continuously updated
- Non-improvable theories are absolute, non-refutable beliefs:
  - Totalitarian ideologies, absolute religions, conspiracy theories, etc.

Fallibility and validity of scientific theories

- All scientific theories are **fallible**
  - May turn out to be false
  - I.e. they are improvable

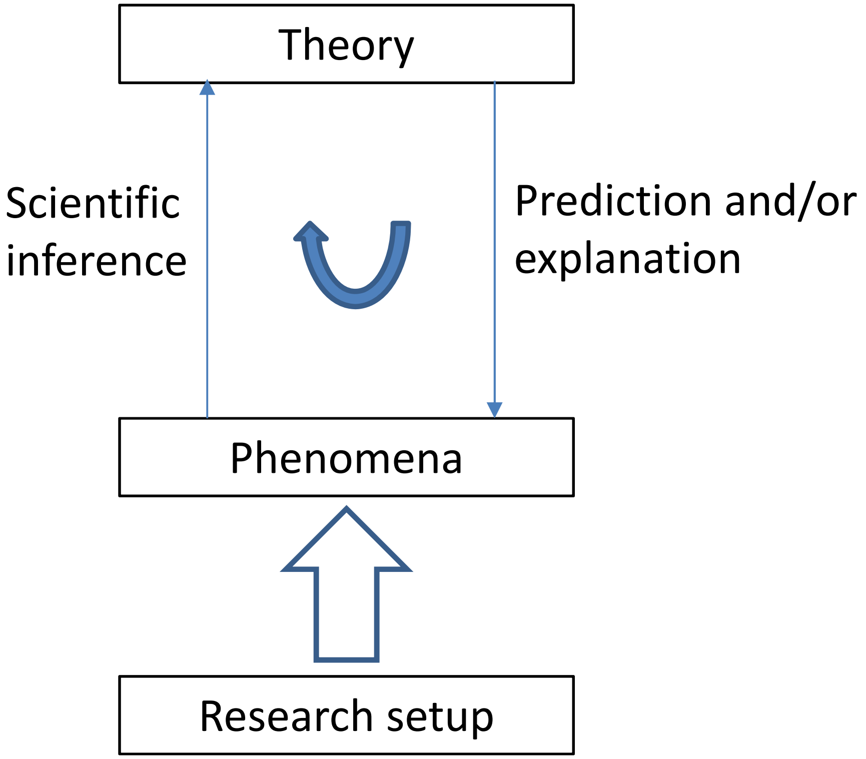
Beliefs in religion, politics, marketing, and social media are usually treated as **infallible** by their defenders, especially if shared by many others.

- **Validity** is degree of support for a belief
  - Degree of (un)certainty must be made explicit in science
  - Never total
  - Outside mathematics there is no certainty
  - No statement about the real world can be "scientifically proven".

Outline

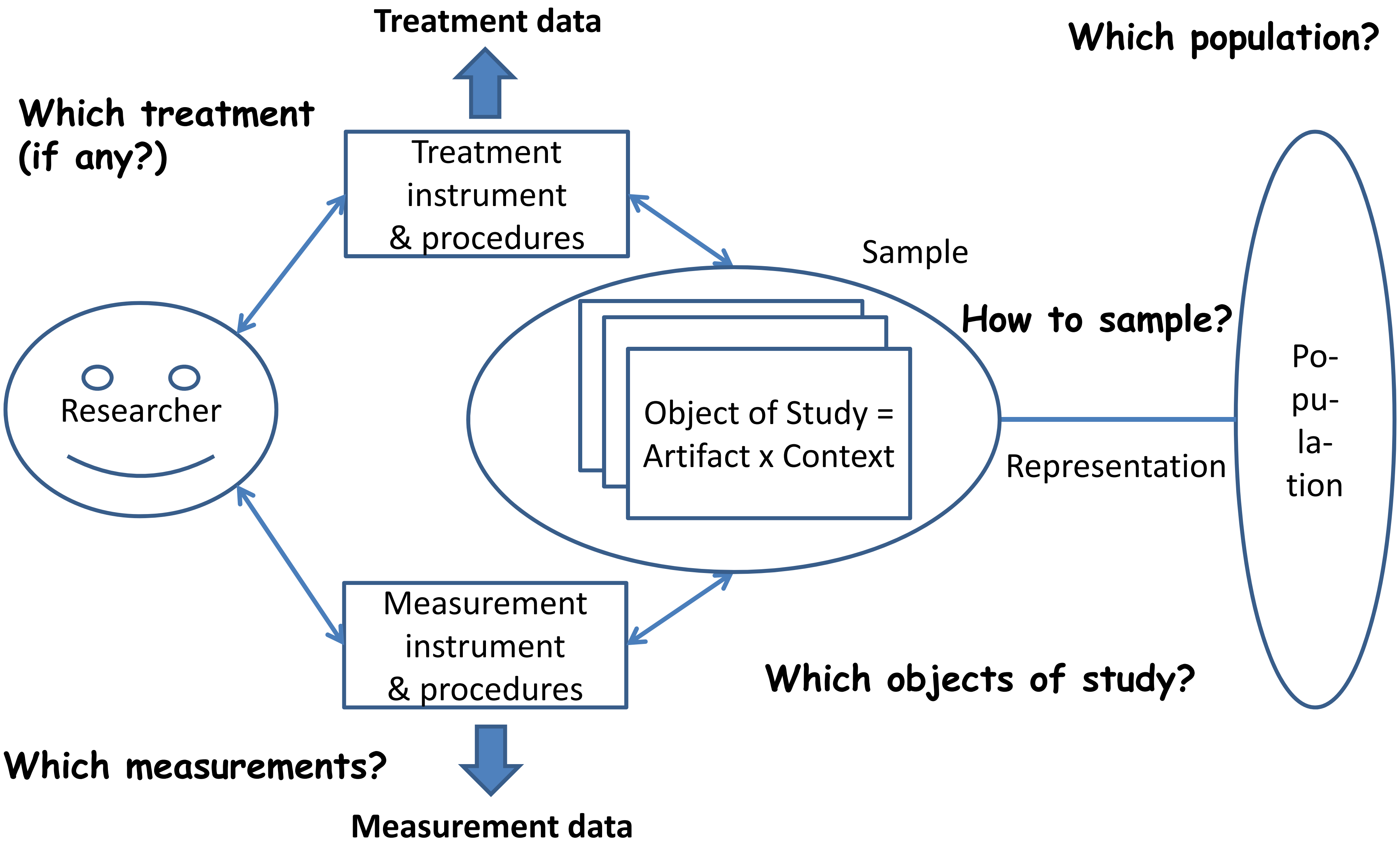
- **Introduction**
  - The design cycle
  - Theories
  - The research setup
- **Scientific inference**
  - Description
  - Explanation
  - Generalization
  - Prediction
- **Empirical research**
  - Checklist
  - Example research methods

Research setup produces phenomena that are measured



© R.J. Wieringa

Design decisions for research setup



Exercise

- C. Hildebrand, G. Häubl, A. Herrmann, J.R. Landwher, When social media can be bad for you: community feedback stifles consumer creativity and reduces satisfaction with self-designed products. Inf. Syst. Res. **24**(1), 14–29 (2013)

Hypothesis 1A.

- *Receiving community feedback on their initial self-design results in assimilation toward the community feedback when consumers choose their final self-designed products.*

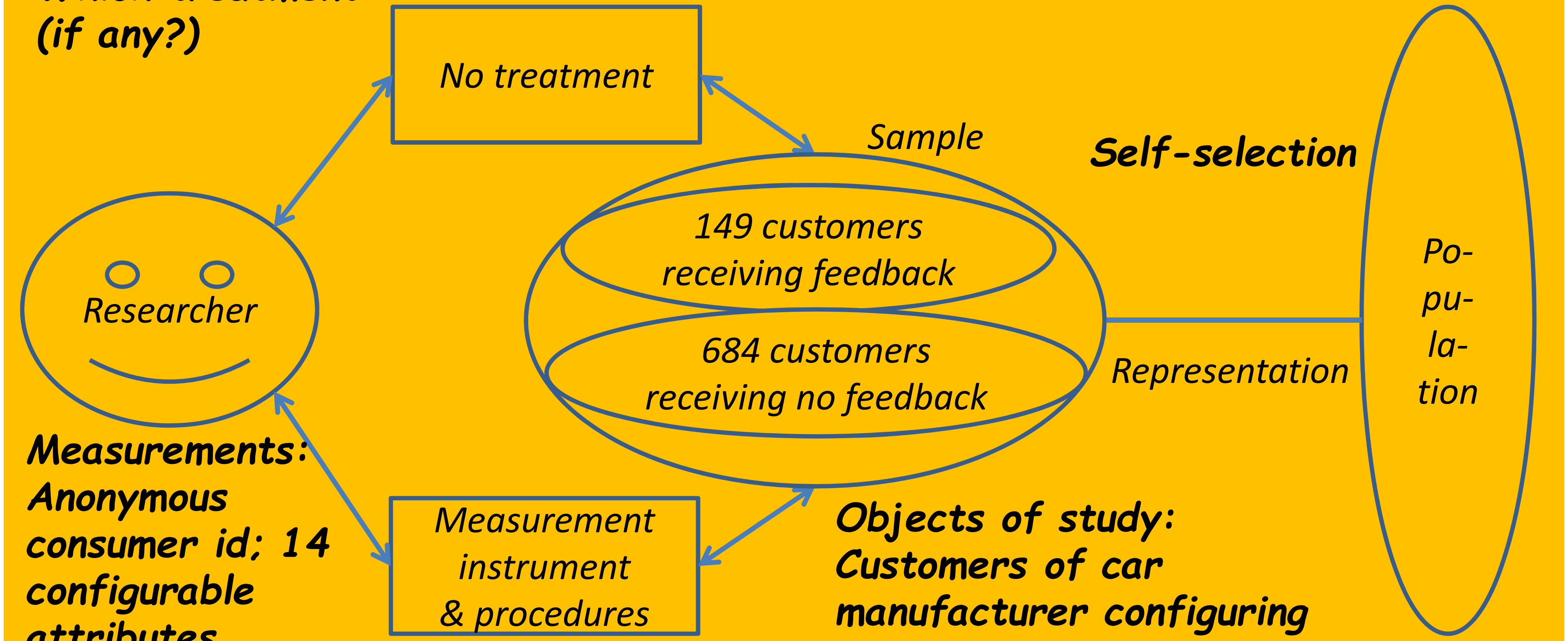
Hypothesis 1B.

- *Assimilation toward the community feedback is stronger when consumers' initial self-designs are more extreme.*

- What is the research setup to test these hypotheses? (see sections 3.1)

Study population: all customers configuring a car. Bigger, theoretical population: all consumers configuring a product

Which treatment (if any?)


No treatment
Sample
Self-selection
149 customers receiving feedback
684 customers receiving no feedback
Researcher
Representation
Po-pu-la-tion
Measurement instrument & procedures

Measurements: Anonymous consumer id; 14 configurable attributes, values for initial and final configuration

Objects of study: Customers of car manufacturer configuring a car

Important kinds of research

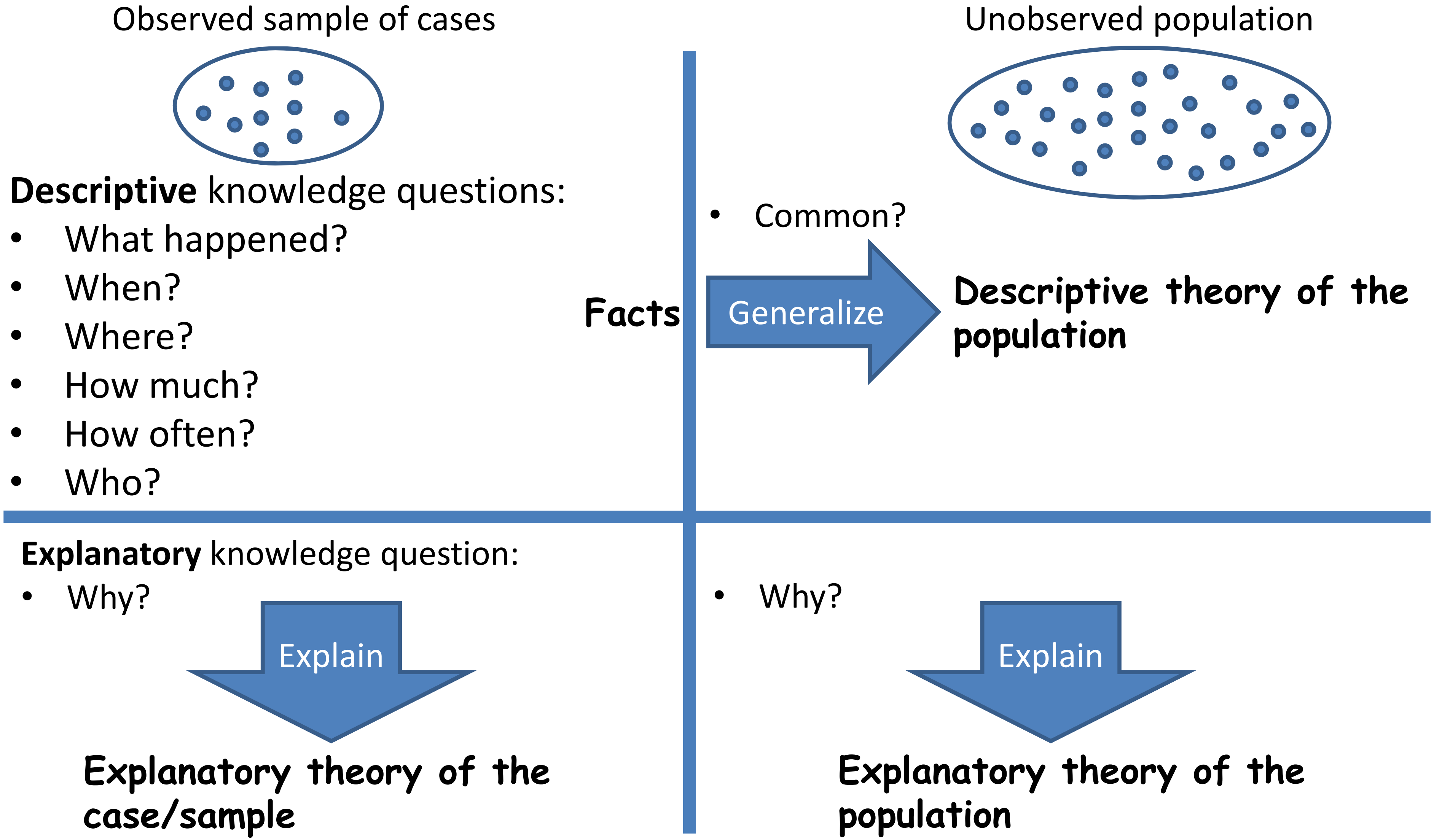
- Case-based versus sample-based setup
- Laboratory versus field (real-world) setup
- Experimental versus observational setup

© R.J. Wieringa

Outline

- **Introduction**
  - The design cycle
  - Theories
  - The research setup
- **Scientific inference**
  - Description
  - Explanation
  - Generalization
  - Prediction
- **Empirical research**
  - Checklist
  - Example research methods

Research questions



Observed sample of cases

Descriptive knowledge questions:

- What happened?
- When?
- Where?
- How much?
- How often?
- Who?

Facts

Unobserved population

- Common?

Generalize

Descriptive theory of the population

Explanatory knowledge question:

- Why?

Explain

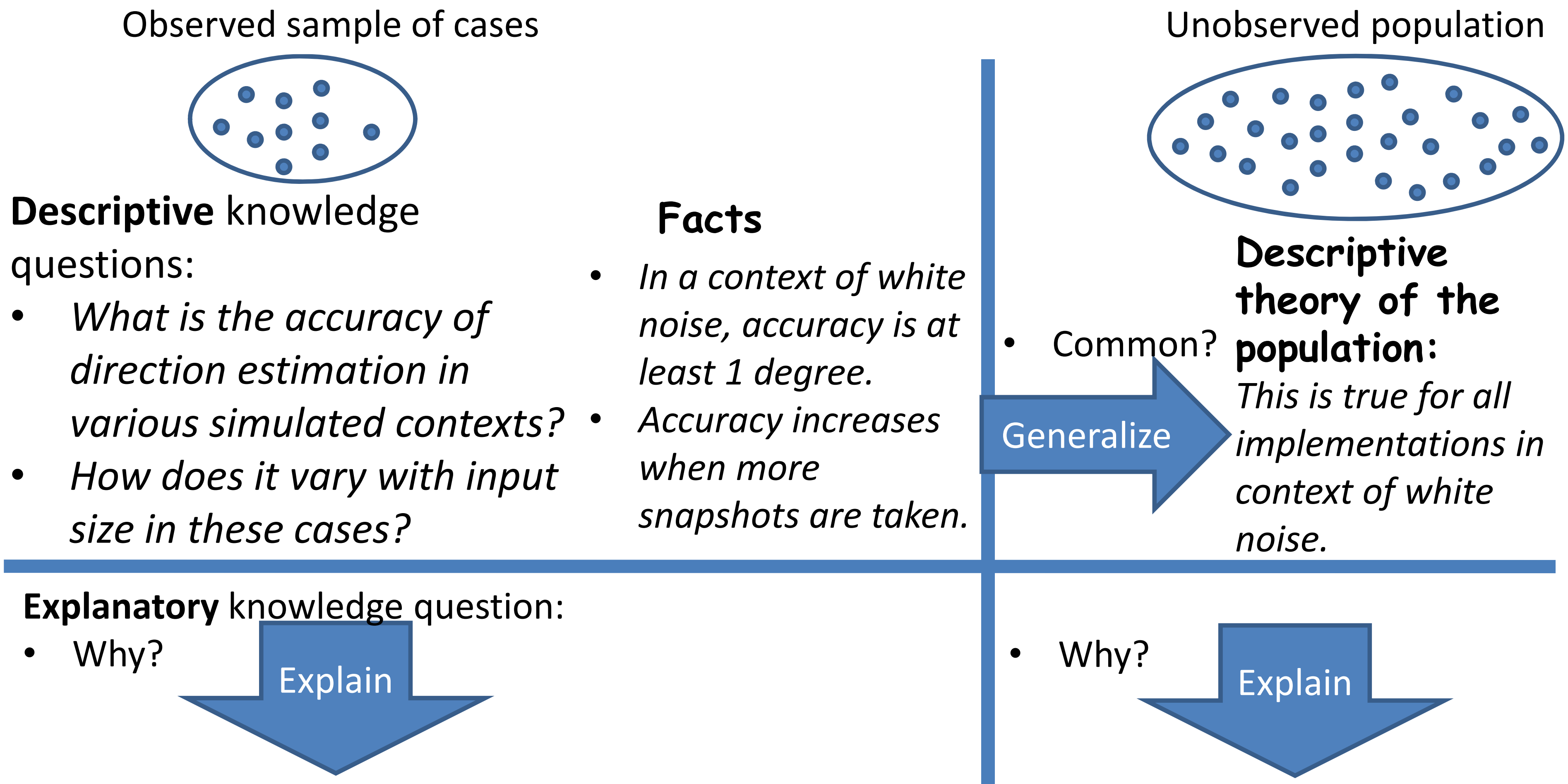
Explanatory theory of the case/sample

- Why?

Explain

Explanatory theory of the population

Questions, factual answers, and theories



Observed sample of cases

Descriptive knowledge questions:

- *What is the accuracy of direction estimation in various simulated contexts?*
- *How does it vary with input size in these cases?*

Facts

- *In a context of white noise, accuracy is at least 1 degree.*
- *Accuracy increases when more snapshots are taken.*

Unobserved population

- Common?

Generalize

Descriptive theory of the population: *This is true for all implementations in context of white noise.*

Explanatory knowledge question:

- Why?

Explain

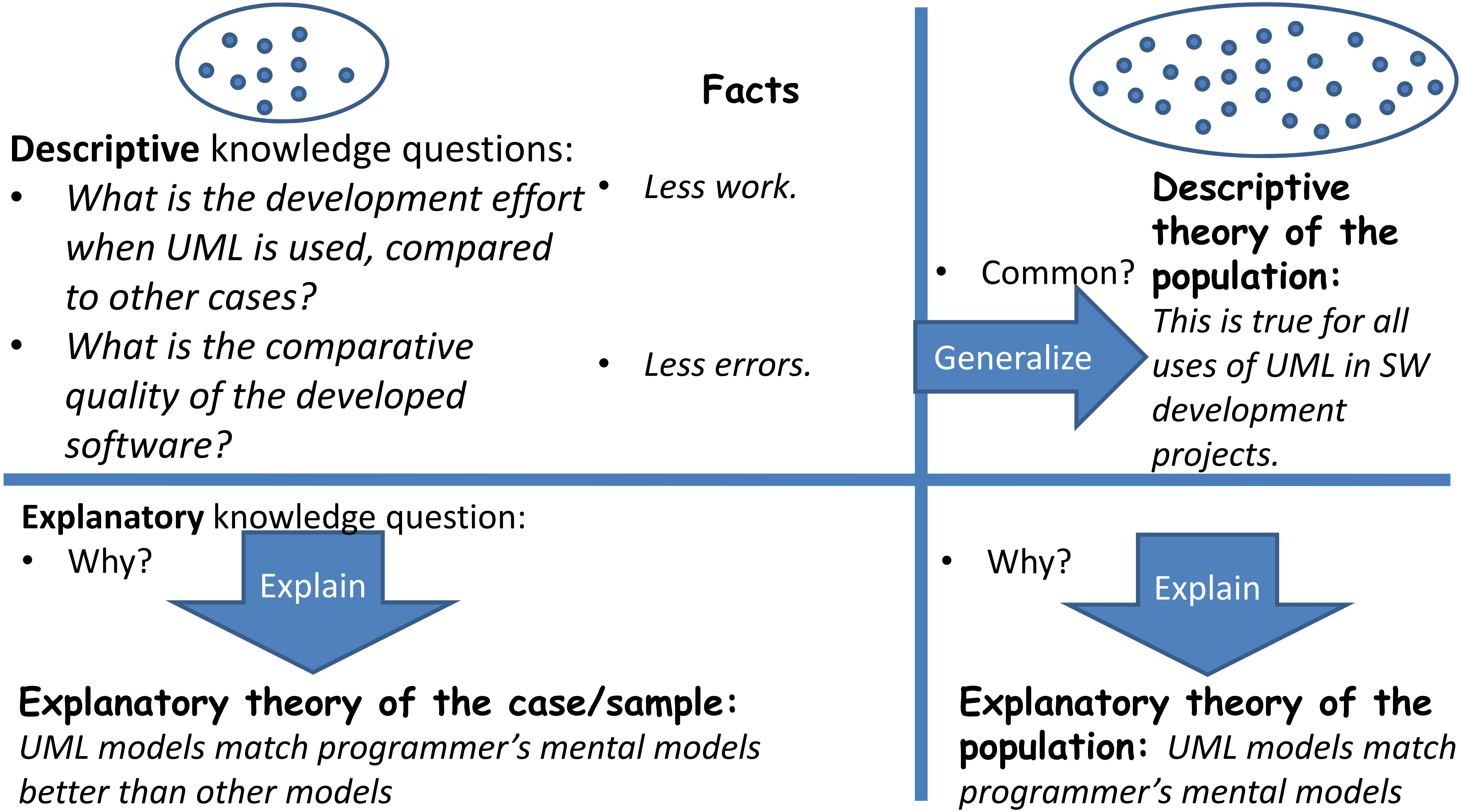
Explanatory theory of the case/sample: *Structure of the algorithm explains output, but not the exact accuracy*

- Why?

Explain

Explanatory theory of the population: *Structure of the algorithm explains output, but not the exact accuracy*

Questions, factual answers, and theories



Observed sample of cases

Unobserved population

Descriptive knowledge questions:

- *What is the development effort when UML is used, compared to other cases?*
- *What is the comparative quality of the developed software?*

Facts

- *Less work.*
- *Less errors.*

- Common?

Generalize

Descriptive theory of the population: *This is true for all uses of UML in SW development projects.*

Explanatory knowledge question:

- Why?

Explain

Explanatory theory of the case/sample: *UML models match programmer's mental models better than other models*

- Why?

Explain

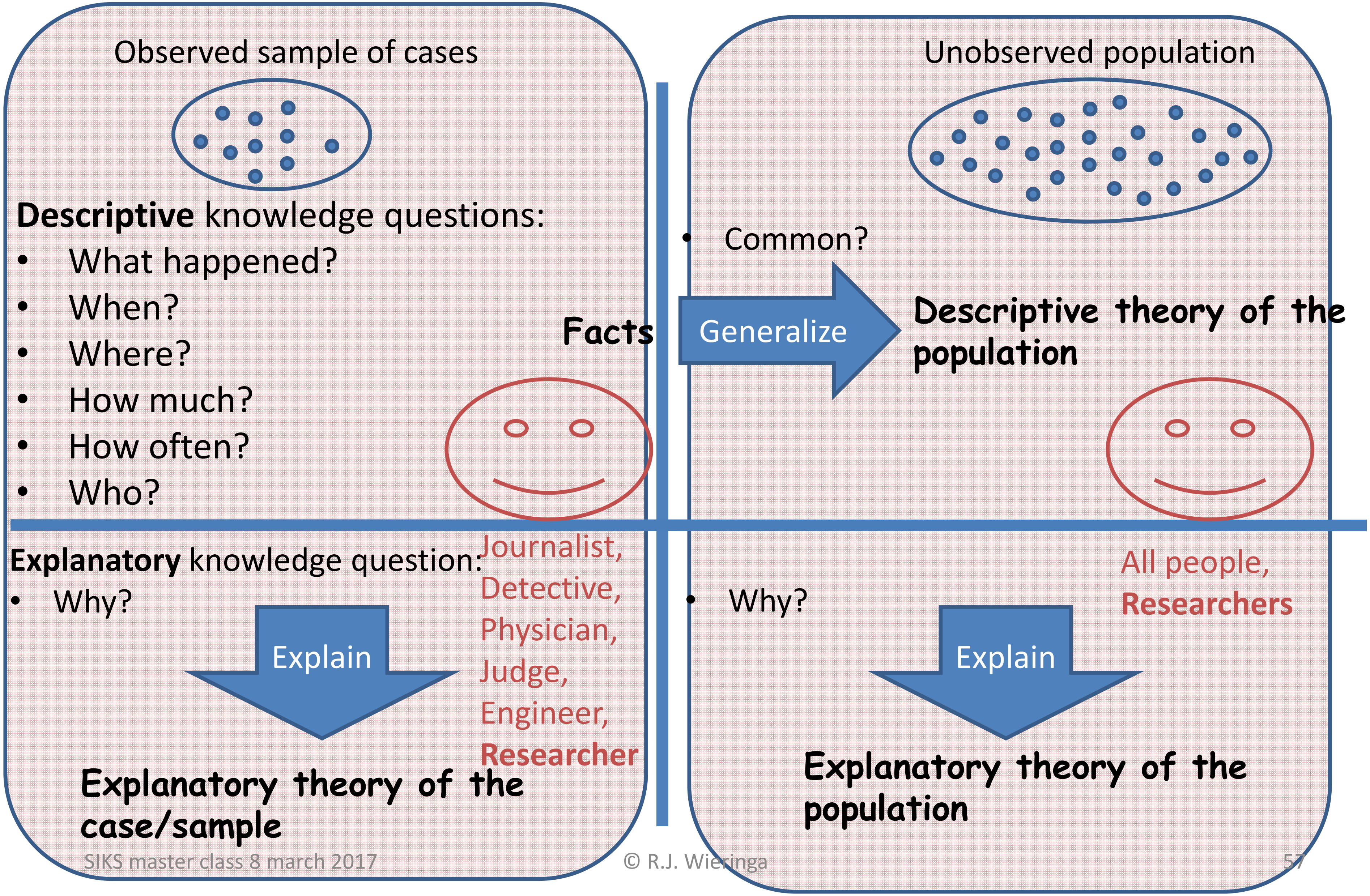
Explanatory theory of the population: *UML models match programmer's mental models better than other models*

Facts

- May be hard to establish.
- In politics, religion, marketing and social media, opinions are treated as facts

- In journalism, crime investigations, medical diagnosis, the court room, engineering, and research, facts should be established beyond <u>reasonable</u> doubt.
  - No opinions
  - No value judgments
  - No ambiguity
- <u>Uncertainty</u> about facts should be acknowledged!
  - Consider the risk (likelihood & impact) of being wrong

Facts, theories, role models



Observed sample of cases

Descriptive knowledge questions:

- What happened?
- When?
- Where?
- How much?
- How often?
- Who?

Facts

Unobserved population

- Common?

Generalize

Descriptive theory of the population

Explanatory knowledge question:

- Why?

Explain

Journalist,
Detective,
Physician,
Judge,
Engineer,
Researcher

Explanatory theory of the case/sample

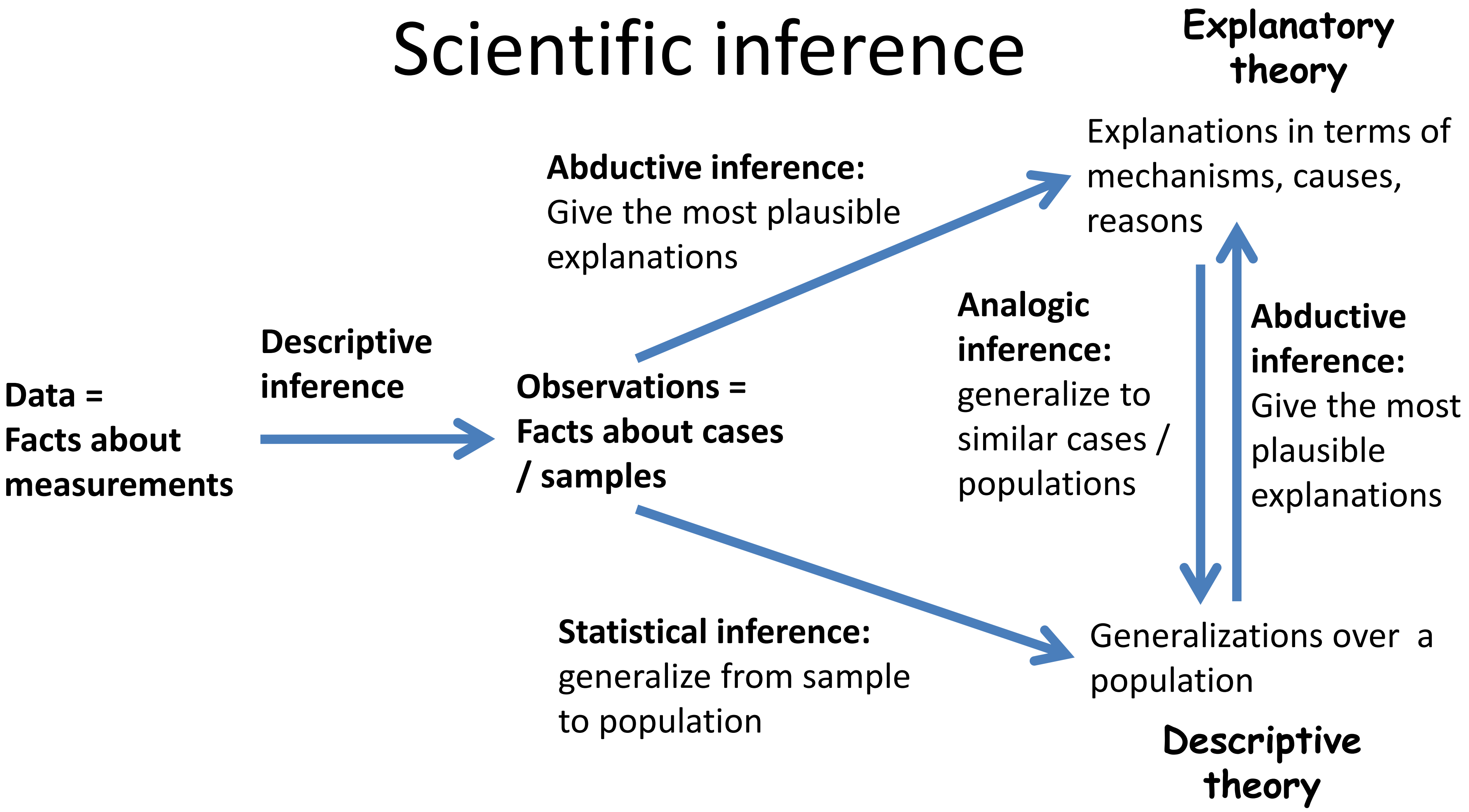
- Why?

Explain

All people,
Researchers

Explanatory theory of the population

From facts to theories: Scientific inference



Explanatory theory

Explanations in terms of mechanisms, causes, reasons

Abductive inference: Give the most plausible explanations

Data = Facts about measurements

Descriptive inference

Observations = Facts about cases / samples

Analogic inference: generalize to similar cases / populations

Abductive inference: Give the most plausible explanations

Statistical inference: generalize from sample to population

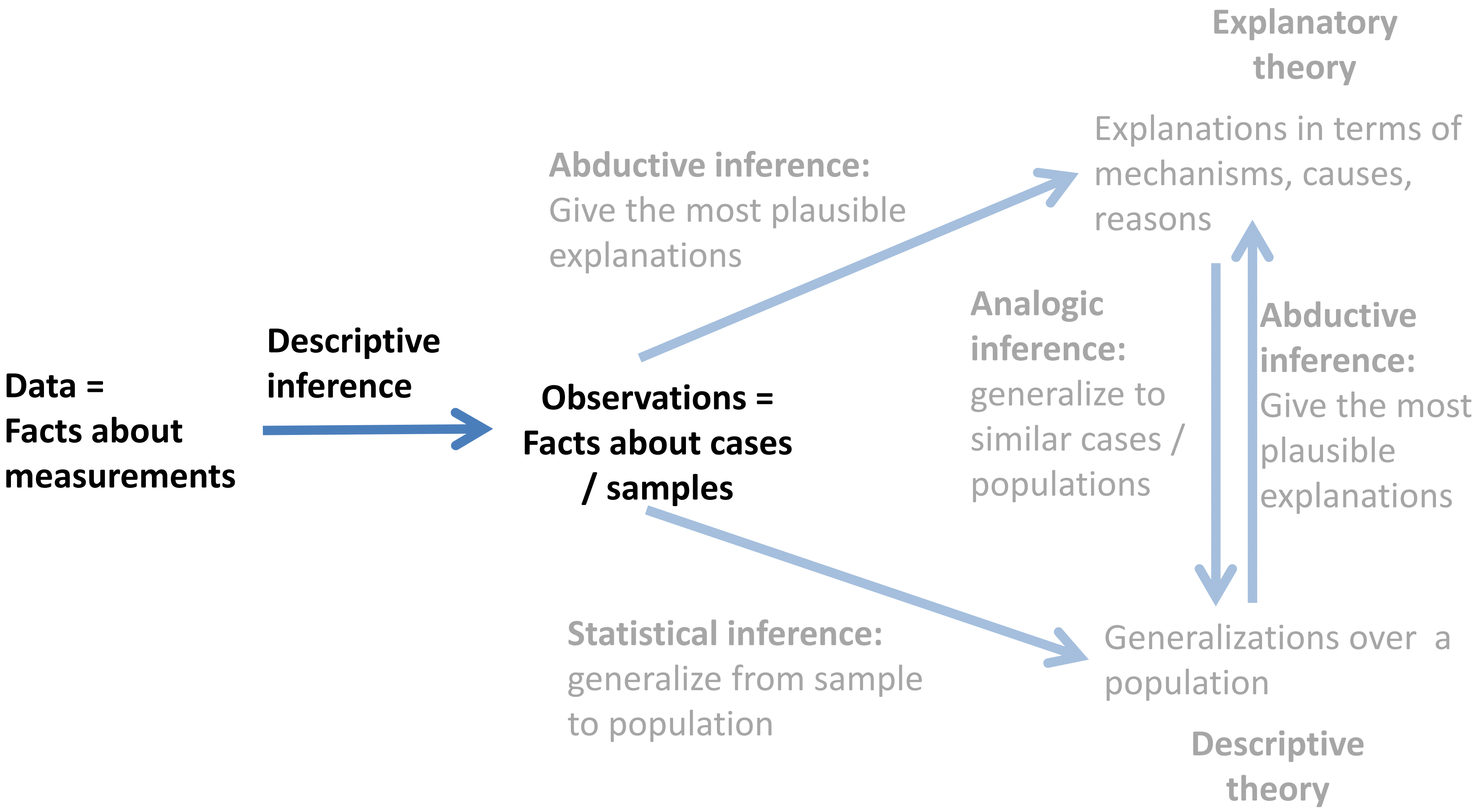
Generalizations over a population

Descriptive theory

Outline

- **Introduction**
  - The design cycle
  - Theories
  - The research setup
- **Scientific inference**
  - Description
  - Explanation
  - Generalization
  - Prediction
- **Empirical research**
  - Checklist
  - Example research methods

From facts to theories



Explanatory theory

Explanations in terms of mechanisms, causes, reasons

Abductive inference: Give the most plausible explanations

Data = **Facts about measurements**

Descriptive inference

Observations = **Facts about cases / samples**

Analogic inference: generalize to similar cases / populations

Abductive inference: Give the most plausible explanations

Statistical inference: generalize from sample to population

Generalizations over a population

Descriptive theory

Measurements

- Records
  - Video recordings
  - Sound recordings
  - Sensor data: temperature, time, position, ….
  - Software data: logs, source code, databases, performance data, …
- Writings
  - Questionnaire answers
  - Notes by researchers or subjects

- It is a fact that you have these measurements.
  - Symbolic (“interpretative”) data: words or images
  - Qualitative data: nominal or ordinal scale
  - Quantitative data: interval or ratio scale

- **All** measurements need to be interpreted to turn them into facts about the cases that you studied!
  - This is descriptive inference

Descriptive inference (Interpretation of measurements)

- Records
  - Video recordings: *removal of bad recordings, # words uttered, direction of gaze, number of turns in conversation, time spent talking, …*
  - Sound recordings: *removal of bad recordings, interview transcripts, coded interviews (content analysis), grounded theory analysis, …*
  - Sensor data: *removal of bad measurements (outliers), definition of measurement scale (nominal, ordinal, interval, ratio), scale transformation, …*
  - Software data: *removal of bad data, reduction of words to stems, …*
- Writings
  - Questionnaire answers: *removal of bad answers, definition of scales, …*
  - Notes by researchers or subjects: *removal of bad data, coding, ...*

Validity of descriptive inference

- **Descriptive validity** is degree of support for a description
- Checks on data preparation:
  - Do the sanitized data represent the same facts as the raw data?
  - Is data removal defensible beyond doubt?
  - Would your opponents produce the same descriptions from the raw data?
- Checks on data interpretation:
  - Would your peers produce the same interpretation?
  - Do the subjects accept your descriptions as facts?
- Check on statistical variables:
  - Chance model (meaning, measurement, distribution, sampling) defined?
- Ask others to prepare and interpret data independently from you.

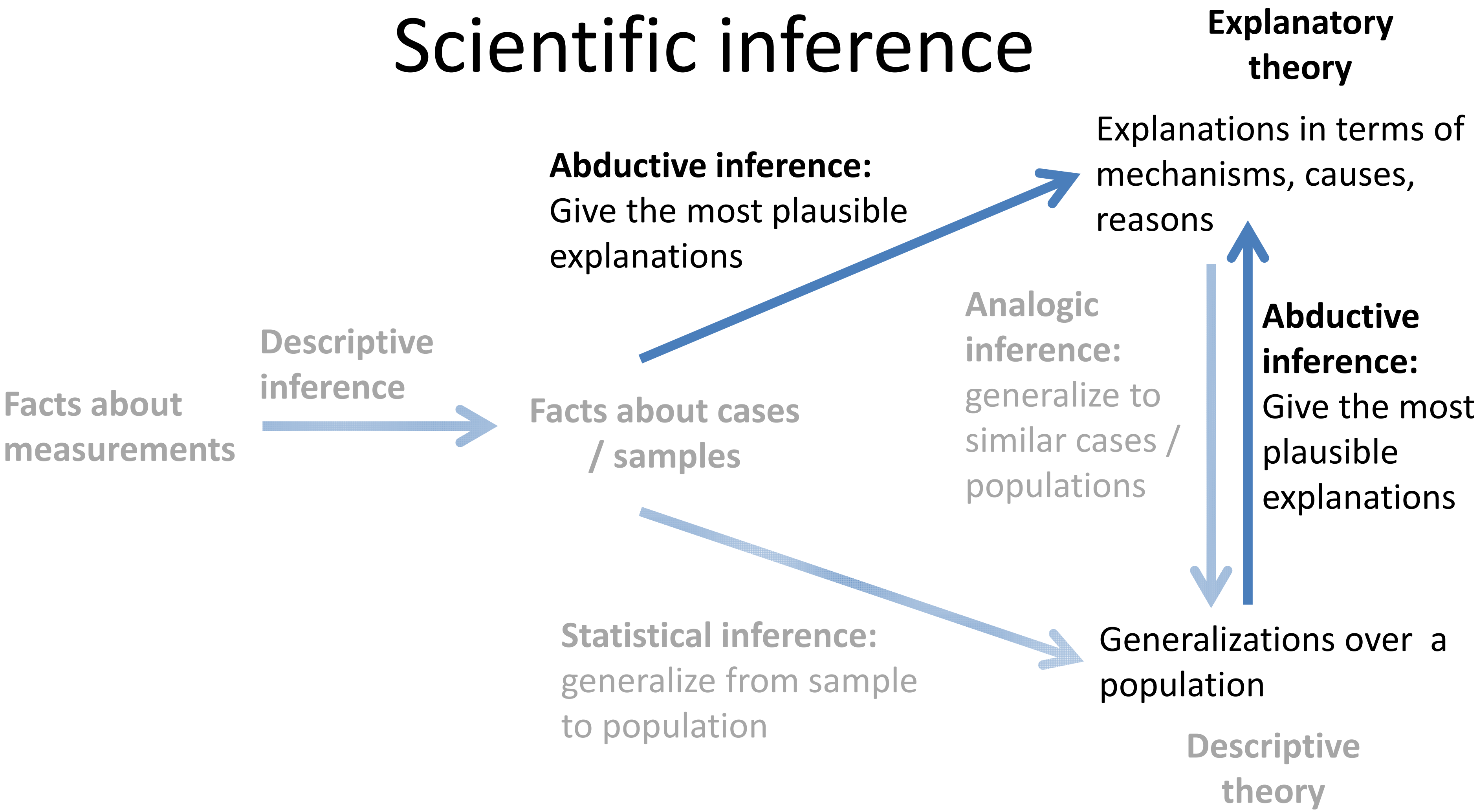
Exercise

- Identify descriptive inference and descriptive validity in
  - C. Hildebrand, G. Häubl, A. Herrmann, J.R. Landwher, When social media can be bad for you: community feedback stifles consumer creativity and reduces satisfaction with self-designed products. Inf. Syst. Res. **24**(1), 14–29 (2013)
- *Transformation of attribute data into Euclidian distances between initial and final configuration: no information added*
- *Weighting the attribute changes by "importance", which is measured by the amount of money consumers spent on an attribute: this is an interpretation. Some cheap changes may be more important than other cheap changes*

Outline

- **Introduction**
  - The design cycle
  - Theories
  - The research setup
- **Scientific inference**
  - Description
  - Explanation
  - Generalization
  - Prediction
- **Empirical research**
  - Checklist
  - Example research methods

From facts to theories: Scientific inference



Explanatory theory

Explanations in terms of mechanisms, causes, reasons

Abductive inference: Give the most plausible explanations

Descriptive inference

Facts about measurements

Facts about cases / samples

Analogic inference: generalize to similar cases / populations

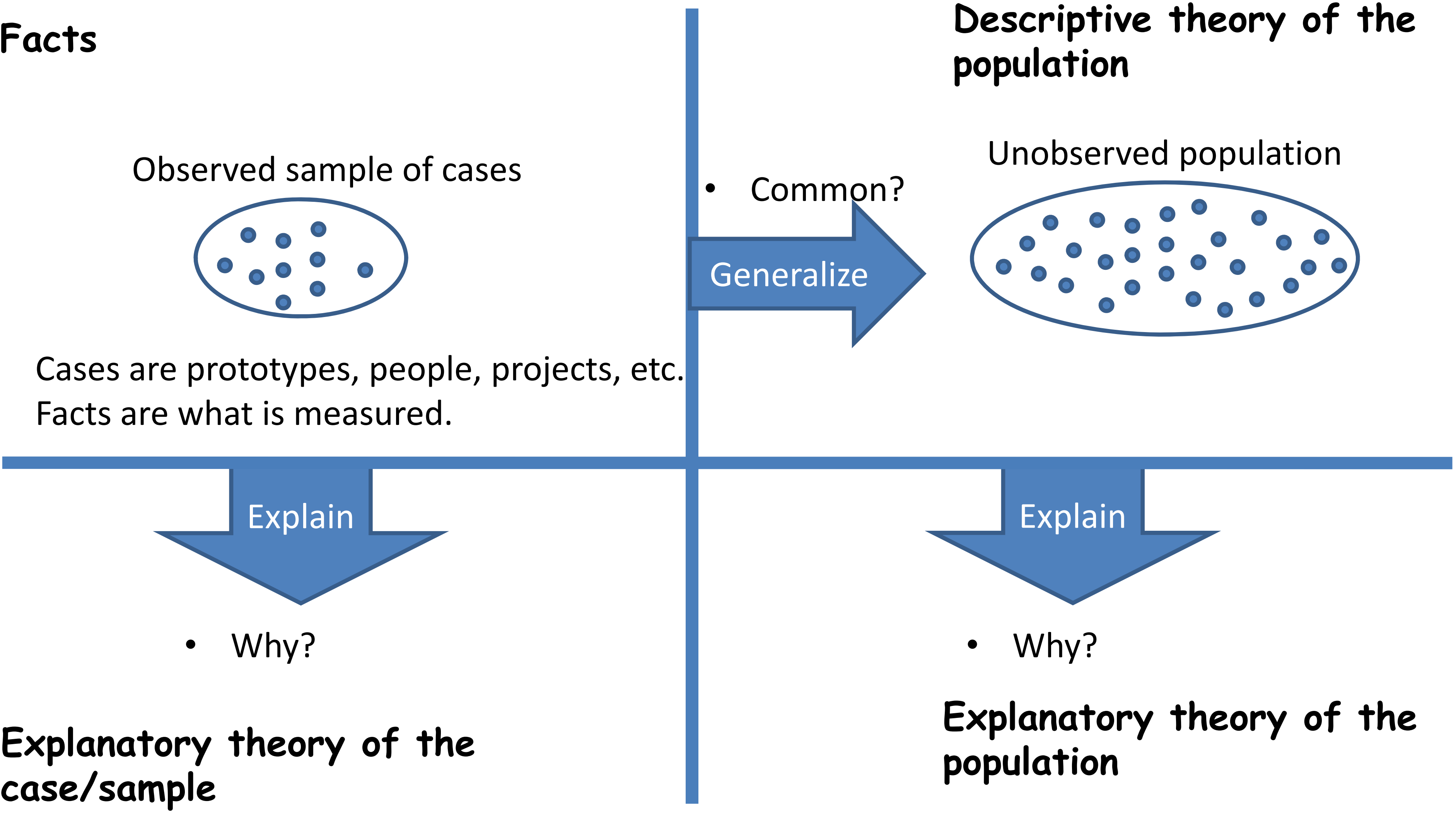
Abductive inference: Give the most plausible explanations

Statistical inference: generalize from sample to population

Generalizations over a population

Descriptive theory

Facts versus theories



Facts

Observed sample of cases

- Cases are prototypes, people, projects, etc.
- Facts are what is measured.

Generalize

- Common?

Descriptive theory of the population

Unobserved population

Explain

- Why?

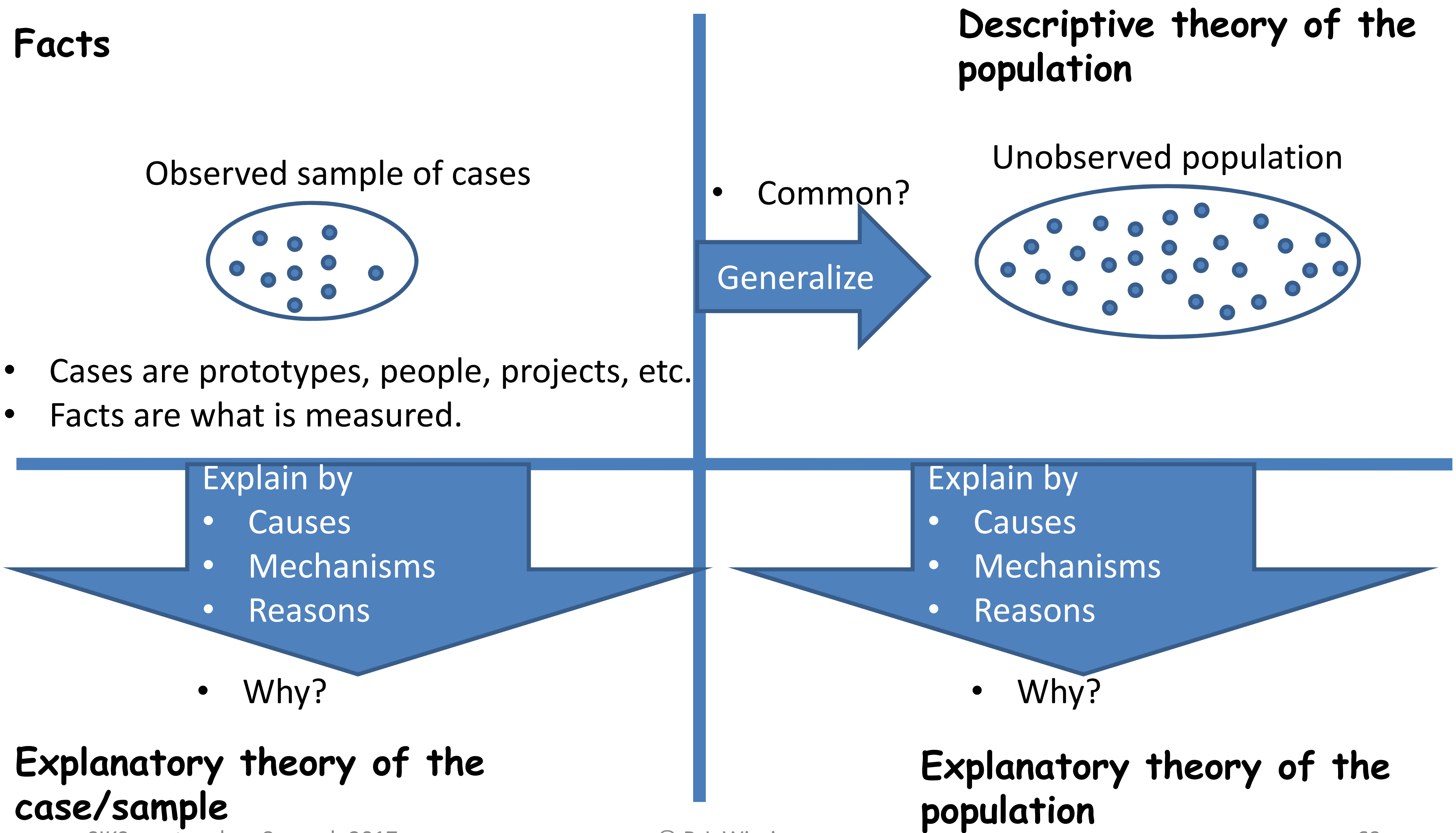
Explanatory theory of the case/sample

Explain

- Why?

Explanatory theory of the population

Three kinds of explanation



Facts

Observed sample of cases

- Cases are prototypes, people, projects, etc.
- Facts are what is measured.

- Common?

Generalize

Descriptive theory of the population

Unobserved population

Explain by
- Causes
- Mechanisms
- Reasons

- Why?

Explanatory theory of the case/sample

Explain by
- Causes
- Mechanisms
- Reasons

- Why?

Explanatory theory of the population

Example 1: light

- *Descriptive question: Is the light on?*
  - *Based on observation: Yes.*
  - *When? Now.*
  - *Where? Here.*
- *Explanatory question: Why is it on?*
  1. ***Cause:** because someone turned the light switch, it is on (and not off).* Explains difference with off-state.
  2. *Why does this cause the light to switch on? **Mechanism:** because the switch and light bulbs are connected by wires to an electricity source, in this architecture …, and these components have these capabilities …..* Explains how on-state is produced.
  3. *By why did someone turn the light on? **Reasons:** Because we wanted sufficient light to be able to read, and it was too dark to read.* Explains which stakeholder goal is contributed to.

Example 2: coffee

- **Causal explanation:** effect attributed to a cause. Explain difference in outcomes by difference in interventions. Causation is difference-making.
  - *The coffee made me stay awake late.*
- **Architectural explanation:** Outcome produced by interaction among components. Explain capability of system in terms of capabilities of components
  - *Mechanism of action: Caffeine has a psychostimulant effect because it antagonizes adenosine, which normally inhibits neurotransmitters such as dopamine.*
- **Rational explanation:** Outcome contributes to a goal. Explain outcome in terms of rational takeholder choices.
  - *I worked late because I wanted to finish the paper before the deadline.*

Example 3: software

- *Descriptive question: What is the performance of this program to estimate direction of arrival of plane waves?*
  - *Execution time for different classes of inputs?*
  - *Memory usage?*
  - *Accuracy?*
  - *Etc. etc.*
- *Explanatory question: Why does this program have this performance (compared to others)?*
  1. ***Cause:*** *Variation in execution time is caused by variation in input; etc.*
  2. ***Mechanism:*** *Execution time varies this way because it has this architecture with these components*
  3. ***Reasons:*** *Observed execution time varies this way because users choose to drive on busy roads with a lot of signal interference*

Example 4: method

- *Descriptive question: What is the performance of this method for developing software?*
  - *Understandability for practioners*
  - *Learnability*
  - *Quality of the result*
  - *Perceived utility*
  - *Etc. etc.*
- *Explanatory question: Why does this method have this performance?*
  1. ***Cause:** Difference in understanding of methods by software engineers is attributed to differences in the methods, and not to differences in people, software systems, etc. (cf. testing of a medicine)*
  2. ***Mechanism:** These differences are explained by the structure of the method and/or the structure of cognition. (cf. mechanism of action of a medicine)*
  3. ***Reasons:** Developers are rewarded if they use the method well*

Internal validity of an explanation

- **Internal validity** = degree of support for an explanation
- Three kinds of internal validity
  - Of causal explanations
  - Of architectural explanations
  - Of rational explanations
- Customarily stated in terms of **threats** that decrease support.

Checks of internal validity of causal explanations

- **Ambiguous relationship:** ambiguous covariation, ambiguous temporal ordering, ambiguous spatial connection?
- **Object of Study (OoS) dynamics:** could there be interaction among OoSs? Could there be historical events, maturation, dropout of OoSs?
- **Sampling influence:** could the selection mechanism influence the OoSs? Could there be a regression effect?
- **Treatment control:** what other factors than the treatment could influence the OoSs? The treatment allocation mechanism, the experimental setup, the experimenters and their expectations, the novelty of the treatment, compensation by the researcher, resentment about the allocation?
- **Treatment instrument validity:** do the treatment instruments have the effect on the OoS that you claim they have?
- **Measurement influence:** will measurement influence the OoSs?

Checks of internal validity of architectural explanations

- **Analysis:** the analysis of the architectural model may not support its conclusions with mathematical certainty.
  - Are components fully specified?
  - Are interactions fully specified?
- **Variation:** do the real case components match the architectural components of the model?
  - Are all model components present in the real-world case?
  - Do they have the same capabilities?
- **Abstraction:** does the architectural model abstract from relevant interactions in the real case?
  - Are there interfering mechanisms in the target case, absent from the model?

Checks of the internal validity of rational explanations

- **Goals:** Does the actor have the goals that the explanation says it has? Consistently across actions?
- **Motivation:** Do the goals motivate the actions as much as the explanation says it does? Could the actions be motivated by other goals as well?

Exercise

- Analyze abductive inference in the paper by Hildebrand et al.

Statistical inference:

- Linear regression of change in subject preferences against the feedback they received. The slope of the line was positive, meaning that subjects’ final preference is closer to the feedback that they received than their initial preference.

Abductive inference: Three possible explanations:

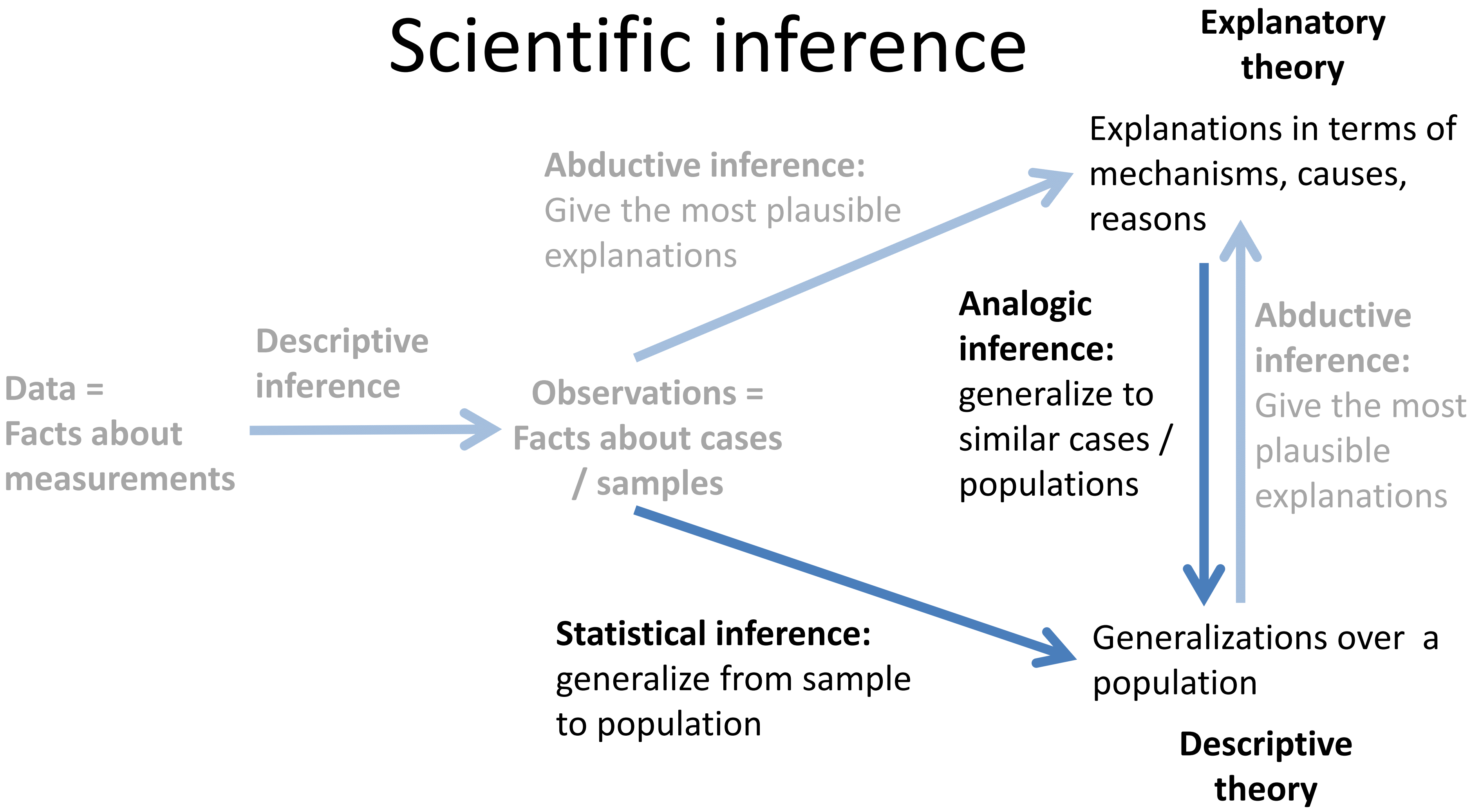
- Subjects receiving feedback seek approval of others. This mechanisms suggests that receiving feedback **causes** preferences to change in the direction of feedback.
- Subjects who self-selected into the treatment would have been more susceptible to the influence of others
- Perhaps subjects shared other characteristics that can explain the observation, such as their age, sex, or education level.

These explanations can be true at the same time! More information is needed to assess their plausibility

Outline

- **Introduction**
  - The design cycle
  - Theories
  - The research setup
- **Scientific inference**
  - Description
  - Explanation
  - Generalization
  - Prediction
- **Empirical research**
  - Checklist
  - Example research methods

From facts to theories: Scientific inference



Explanatory theory

Explanations in terms of mechanisms, causes, reasons

Abductive inference: Give the most plausible explanations

Descriptive inference

Data = Facts about measurements

Observations = Facts about cases / samples

Analogic inference: generalize to similar cases / populations

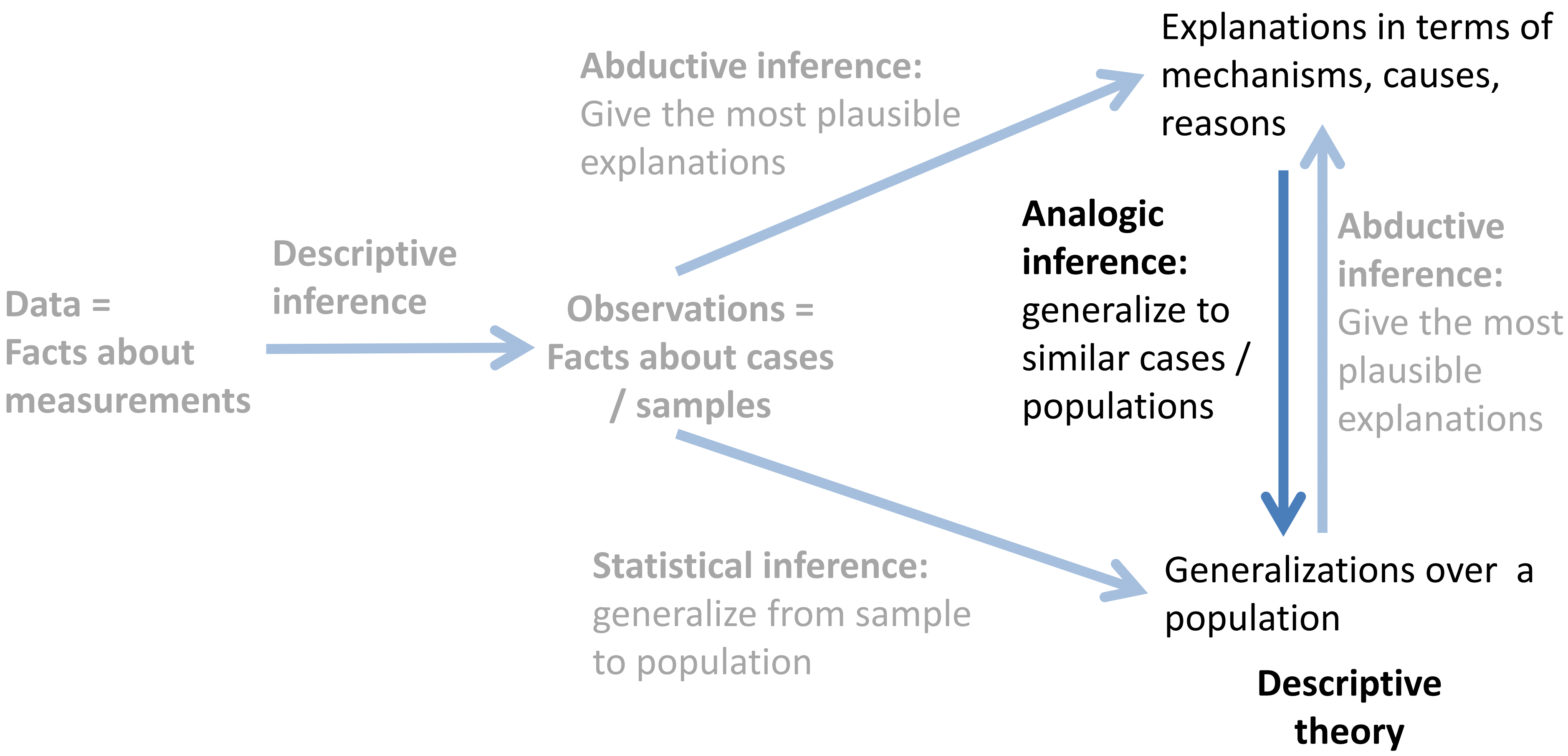
Abductive inference: Give the most plausible explanations

Statistical inference: generalize from sample to population

Generalizations over a population

Descriptive theory

From facts to theories: Scientific inference



Explanatory theory

Explanations in terms of mechanisms, causes, reasons

Abductive inference: Give the most plausible explanations

Data = **Facts about measurements**

Descriptive inference

Observations = **Facts about cases / samples**

Analogic inference: generalize to similar cases / populations

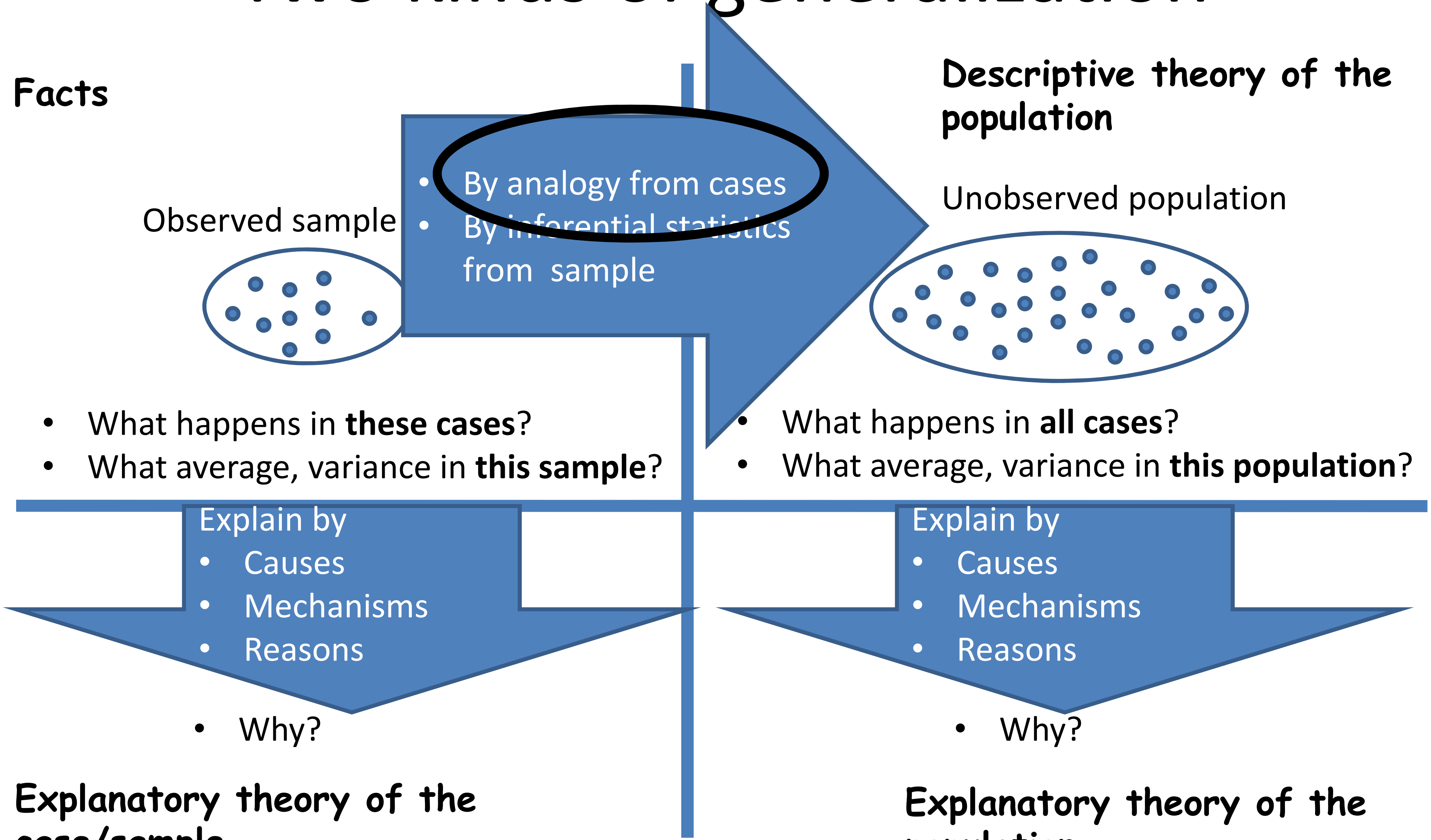
Abductive inference: Give the most plausible explanations

Statistical inference: generalize from sample to population

Generalizations over a population

Descriptive theory

Two kinds of generalization



Facts

Observed sample

- By analogy from cases
- By inferential statistics from sample

Descriptive theory of the population

Unobserved population

- What happens in **these cases**?
- What average, variance in **this sample**?

- What happens in **all cases**?
- What average, variance in **this population**?

Explain by
- Causes
- Mechanisms
- Reasons

- Why?

Explain by
- Causes
- Mechanisms
- Reasons

- Why?

Explanatory theory of the case/sample

Explanatory theory of the population

Case-based generalization

- Precedes sample-based generalization in history.
- Examples
  - *Newton's prism experiment*
  - *Pascal's experiment with air pressure*
  - *Lavoisier's experiments with phosphorus*
  - *Oersted's experiment with magnetic needle*
- "If you build the same research setup, it will exhibit the same phenomena."
  - Similarity
  - Architecture (components and interactions)
  - Repeatability

Generalization by analogy: example

- *Observation:*
  - *Artifact: This prototype implementation of the MUSIC algorithm,*
  - *Context: when used to recognize direction of arrival of plane waves received by an antenna array, in the presence of only white noise, running on a Montium 2 processor,*
  - *Effect: has execution speed less than 7.2 ms and accuracy of at least 1 degree.*
- *Explanations:*
  - *algorithm theory and signal theory*
- *Generalization by analogy:*
  - *All* ***similar*** *implementations*
  - *Running in* ***simila****r contexts*
  - *Will show* ***similar*** *performance* …. always??

…. unless

- *The components in the target case have different capabilities from those in the source cases, or*
- *There are interfering mechanisms in the target case, not present in the source architecture*

Generalization by analogy: example

- *Observations:*
  - *Artifact: This version of the UML*
  - *Context: Used in this software project*
  - *Effect: Produces software with less errors and less effort than in similar projects without the UML,*
- *Explanation:*
  - *UML models are easier to understand for software engineers because they resemble the domain more than other kinds of models, and so the software engineers make less errors and there is less rework.*
- *Generalization*
  - *In similar projects,*
  - *UML*
  - *will have similar effects*

…. unless

- *The tools or actors in the target case have different capabilities from those in the source cases, or*
- *There are interfering mechanisms in the target case, not present in the source architecture, such as political power struggles or high personnel turnover*

Generalization by analogy: general pattern

- All artifacts with similar architecture
- Used in contexts with similar architecture
- Will show similar effects
- Unless
  - the target case **components** have different capabilities than the source cases, or
  - the target case has a different **interactions** than the source cases

Analogic generalization must be supported by an architectural explanation

- “In general, components with these capabilities, in this architecture, will produce this phenomenon”

- Nonexample:
  - *Wallnuts look like brains.*
  - *Brains can think.*
  - *Therefore .... wallnuts can think*
- This is only superficial similarity
  - There is no mechanism that produces thinking in brains and wallnuts!

Generalization by analogy (1)

- *Observation:*
  - *Artifact: A light switch*
  - *Context: next to the door in the wall of a room with ceiling lights*
  - *Effect: toggles the ceiling light on and off.*
- *Explanation:*
  - *The switch and context architectures produce this behavior*
- *Generalization by analogy:*
  - *All **similar** switches*
  - *Running in **simila**r contexts*
  - *Will show **similar** effects*

Descriptive generalization. Implicit assumptions:

1. The mechanisms that explain this performance will be present in all similar artifacts and contexts, and
2. will not be undone by other mechanisms.

Generalization by analogy (2)

- *Observation:*
  - *Artifact: This prototype implementation of the MUSIC algorithm,*
  - *Context: when used to recognize direction of arrival of plane waves received by an antenna array, in the presence of only white noise, running on a Montium 2 processor,*
  - *Effect: has execution speed less than 7.2 ms and accuracy of at least 1 degree.*
- *Explanation:*
  - *Algorithm structure*
- *Generalization by analogy:*
  - *All **similar** implementations*
  - *Running in **simila**r contexts*
  - *Will show **similar** performance*

Descriptive generalization. Implicit assumptions:

1. The mechanisms that explain this performance will be present in all similar artifacts and contexts, and
2. will not be undone by other mechanisms.

Generalization by analogy (3)

- *Observations:*
  - *Artifact: this version of the UML*
  - *Context: Used in this software project*
  - *Effect: Produces software with less errors and less effort than in similar projects without the UML,*
- *Explanation:*
  - *UML models are easier to understand for software engineers because they resemble the domain more than other kinds of models,*
  - *so the software engineers make less errors and there is less rework.*
- *Generalization*
  - *In similar projects, UML will have similar effects*
  - *Assumptions: The mechanisms that produced these effects will be present in all similar projects, i.e. UML is used in the same way, and any relevant social and cognitive mechanisms are present in similar projects too, and*
  - *The effects will not be undone by other mechanisms*

Generalization by analogy

- Must be based on architectural similarity
  - Similar components, with similar capabilities
  - Similar mechanisms involving these components

- Analogy based in similarity of superficial features, without knowledge of underlying mechanisms, is too weak a basis for generalization.
  - *Wallnuts and brains do not share the mechanism that produces thinking in brains*

Two kinds of generalization, so two kinds of validity

- **External validity** = Degree of support for generalization by analogy
- **Conclusion validity** = Degree of support for a sample-based generalization

External validity of analogic generalization

- External validity of analogic generalizations depends on validity of architectural explanation in the target case
  - **Variation:** do the target case components match the architectural components of the model?
    - Are all model components present in the real-world case?
    - Do they have the same capabilities?
  - **Abstraction:** does the architectural model abstract from relevant interactions in the target case?
    - Are there interfering mechanisms in the target case, absent from the model?
- Next slides list mechanisms in the research setup that decrease external validity

Checks on external validity

- Object of study
  - **Similarity:** Does the OoS satisfy the population predicate?
  - **Ambiguity:** Does the OoS satisfy other population predicates too?
- Representative sampling
  - **Case-based research:** Selected cases representative of the population?
- Treatment
  - **Treatment similarity:** Experimental treatment similar to real treatments?
  - **Compliance:** Is the treatment performed as specified?
  - **Treatment control:** Other factors that could influence the OoSs?
- Measurement
  - **Construct validity:** are the definitions of constructs to be measured valid?
  - **Measurement instrument validity?**
  - **Construct levels:** Representative measured range of values?

Analytic induction

- External validity of generalizations can be tested and improved.
- Analytic induction:
  1. Start with an initial theory about how mechanisms produce phenomena
  2. Select a **confirming** or **falsifying** case
  3. Do case study
  4. Update the theory (conceptual framework and/or generalization)
  5. Stop when budget is finished or theory appears stable
- This may give us a theory of similitude:
  – Theory about how similarities and differences between source and target allow prediction of properties in target.

Exercise

Analyze analogic inference in the paper by Hildebrand et al.

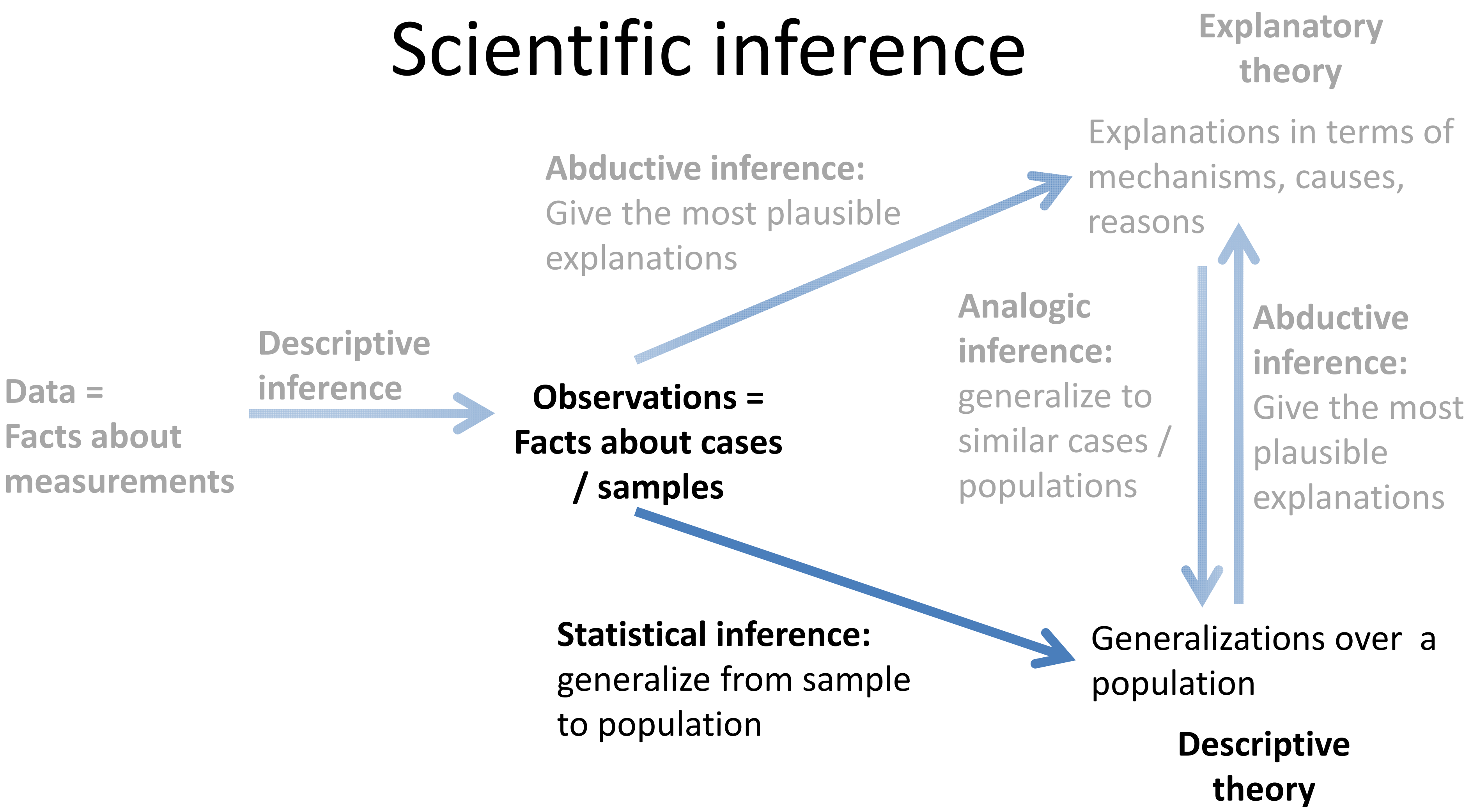
- *Subjects receiving feedback changed their preferences in the direction of feedback.*
- *Possible mechanism that explains this: people tend to conform to the opinion of peers*
- *This supports the claim the receiving feedback causes preferences to changes in the direction of feedback.*
- *To the extent that the mechanism is general, this will happen in other people too.*

- Analogic generalization must be based on architectural explanation

Outline

- **Introduction**
  - The design cycle
  - Theories
  - The research setup
- **Scientific inference**
  - Description
  - Explanation
  - Generalization
    - Case-based
    - Sample-based
  - Prediction
- **Empirical research**
  - Checklist
  - Example research methods

From facts to theories: Scientific inference



Explanatory theory

Explanations in terms of mechanisms, causes, reasons

Abductive inference: Give the most plausible explanations

Data = Facts about measurements

Descriptive inference

Observations = Facts about cases / samples

Analogic inference: generalize to similar cases / populations

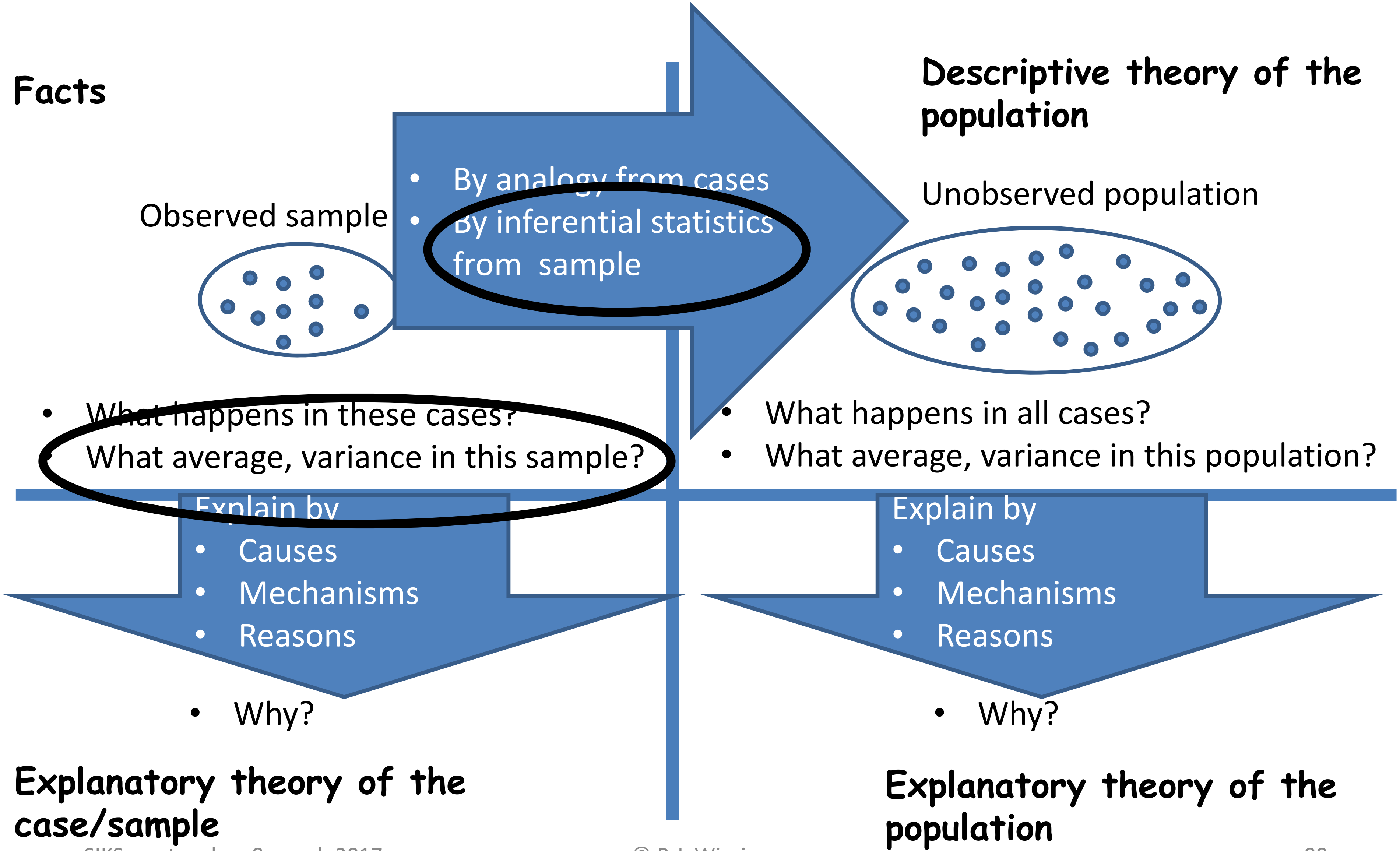
Abductive inference: Give the most plausible explanations

Statistical inference: generalize from sample to population

Generalizations over a population

Descriptive theory

Two kinds of generalization



Facts

Observed sample

- By analogy from cases
- By inferential statistics from sample

Descriptive theory of the population

Unobserved population

- What happens in these cases?
- What average, variance in this sample?

- What happens in all cases?
- What average, variance in this population?

Explain by
- Causes
- Mechanisms
- Reasons

Explain by
- Causes
- Mechanisms
- Reasons

- Why?

- Why?

Explanatory theory of the case/sample

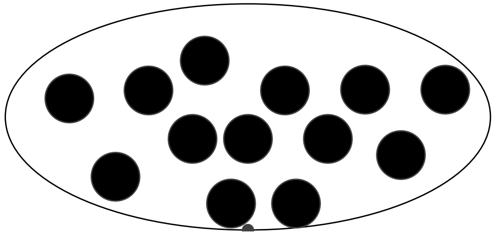
Explanatory theory of the population

Descriptive statistics

- Summarize information in a sample
  - Sample mean, median, mode
  - Sample variance, standard deviation, max, min
  - Sample correlation

Methodology of statistical inference

Theoretical population

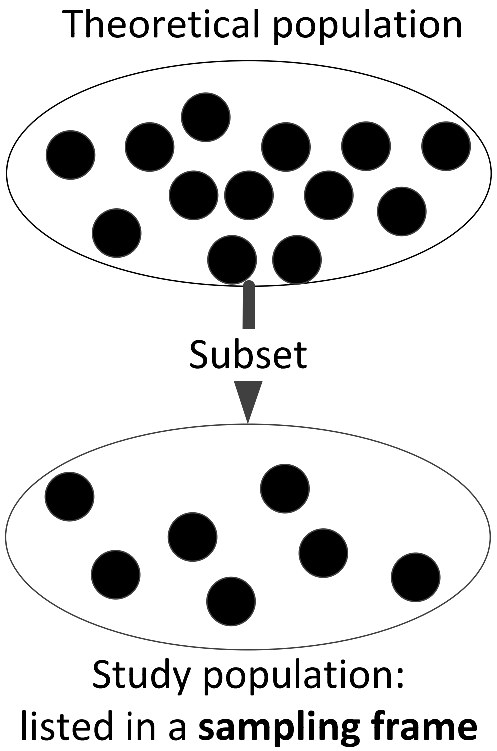


E.g.

- *The set of all instances of an algorithm running in a context;*
- *The set of all global SE projects;*
- *Etc.*

Our ultimate target of generalization

Methodology of statistical inference



Research methodology:

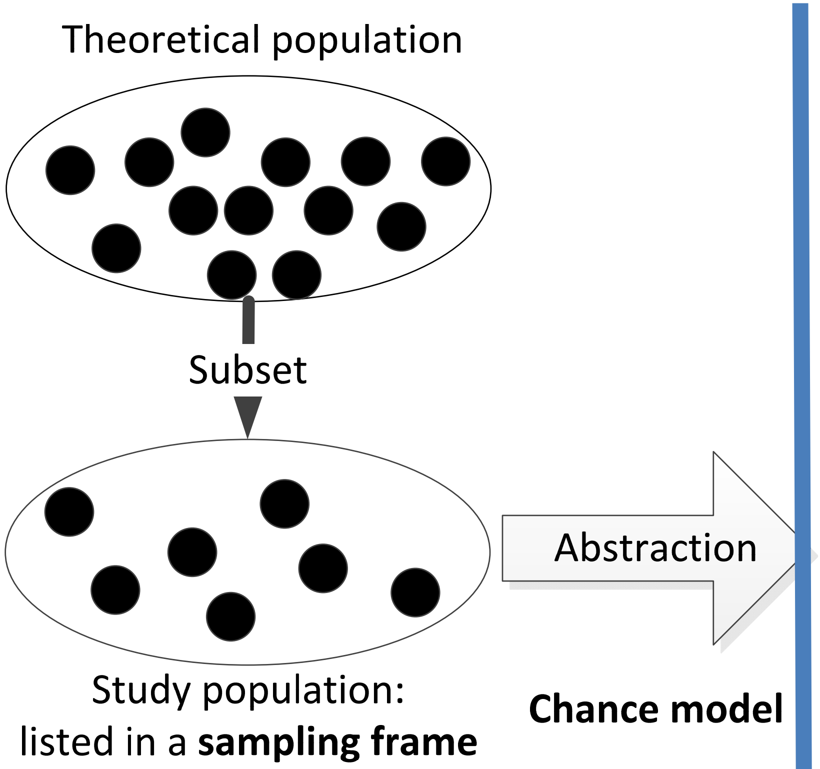
- Sampling frame

E.g.

- *The set of all* ***prototype instances*** *of an algorithm running in a* ***laboratory*** *context;*
- *The set of all global SE projects* ***engaged in by company A****;*
- *Etc.*

The population elements from which you will select a sample

Methodology of statistical inference

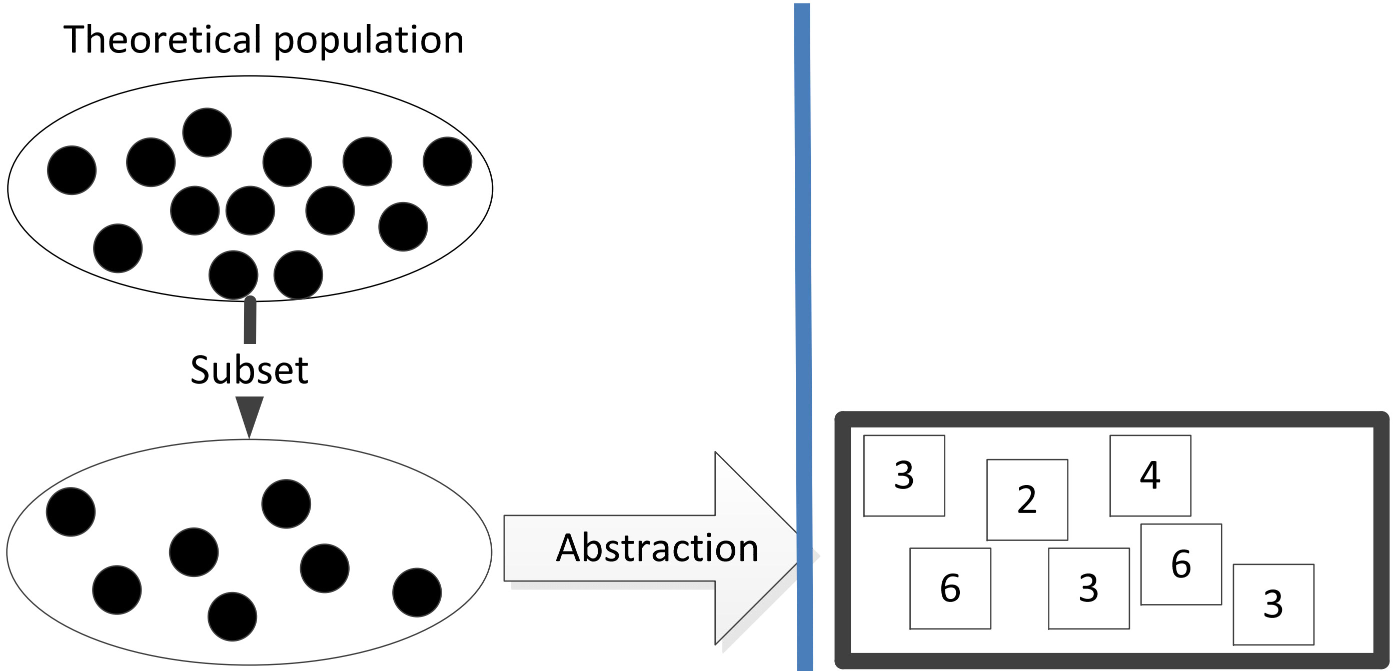


X *The variable that you are interested in*

Research methodology:

- Sampling frame,
- Chance model

Methodology of statistical inference



Theoretical population

Subset

Abstraction

Study population:
listed in a **sampling frame**

Chance model

X-Box: Distribution of X over study population

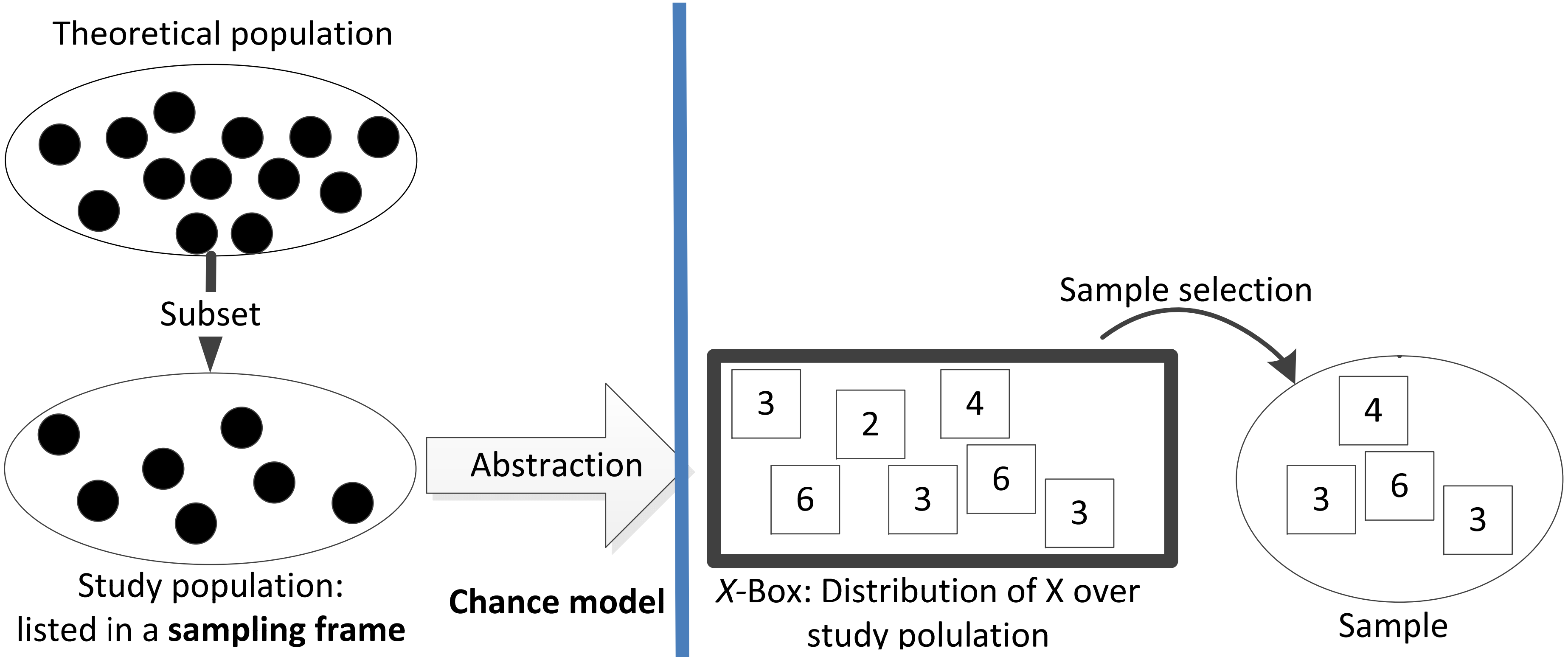
Research methodology.

- Sampling frame,
- Chance model

Statistical inference.

- Unobservable distribution of numbers,

Methodology of statistical inference



Theoretical population

Subset

Study population: listed in a **sampling frame**

Abstraction

Chance model

Sample selection

X-Box: Distribution of X over study polulation

Sample

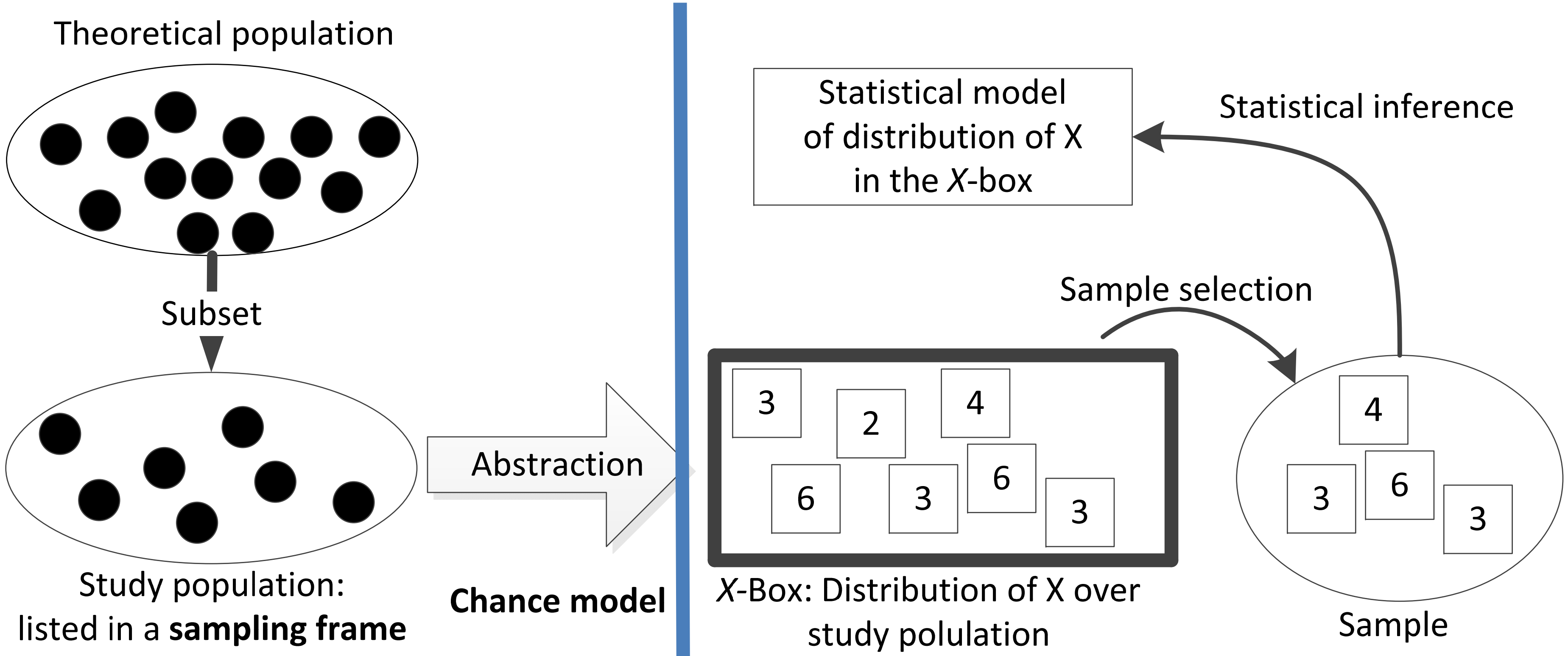
Statistical inference.

- Unobservable distribution of numbers,
- Sample selection,

Research methodology.

- Sampling frame,
- Chance model

Methodology of statistical inference



Theoretical population

Subset

Study population:
listed in a **sampling frame**

Abstraction

Chance model

Statistical model
of distribution of X
in the *X*-box

Statistical inference

Sample selection

X-Box: Distribution of X over
study polulation

Sample

Research methodology.

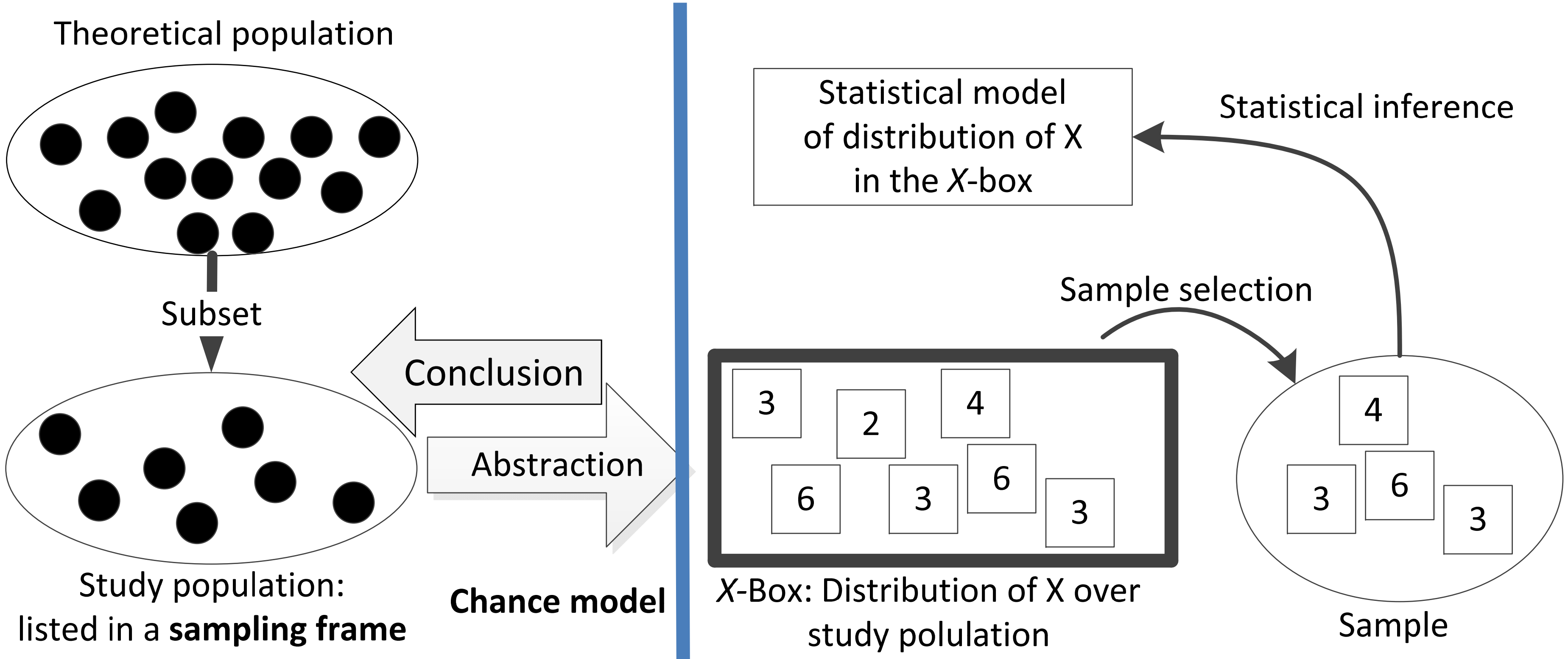
- Sampling frame,
- Chance model

Statistical inference.

- Unobservable distribution of numbers,
- Sample selection,
- Conclusion about unobservable distribution of numbers

© R.J. Wieringa

Methodology of statistical inference



Theoretical population

Subset

Study population: listed in a **sampling frame**

Conclusion

Abstraction

Chance model

Statistical model of distribution of X in the *X*-box

Statistical inference

Sample selection

X-Box: Distribution of X over study polulation

Sample

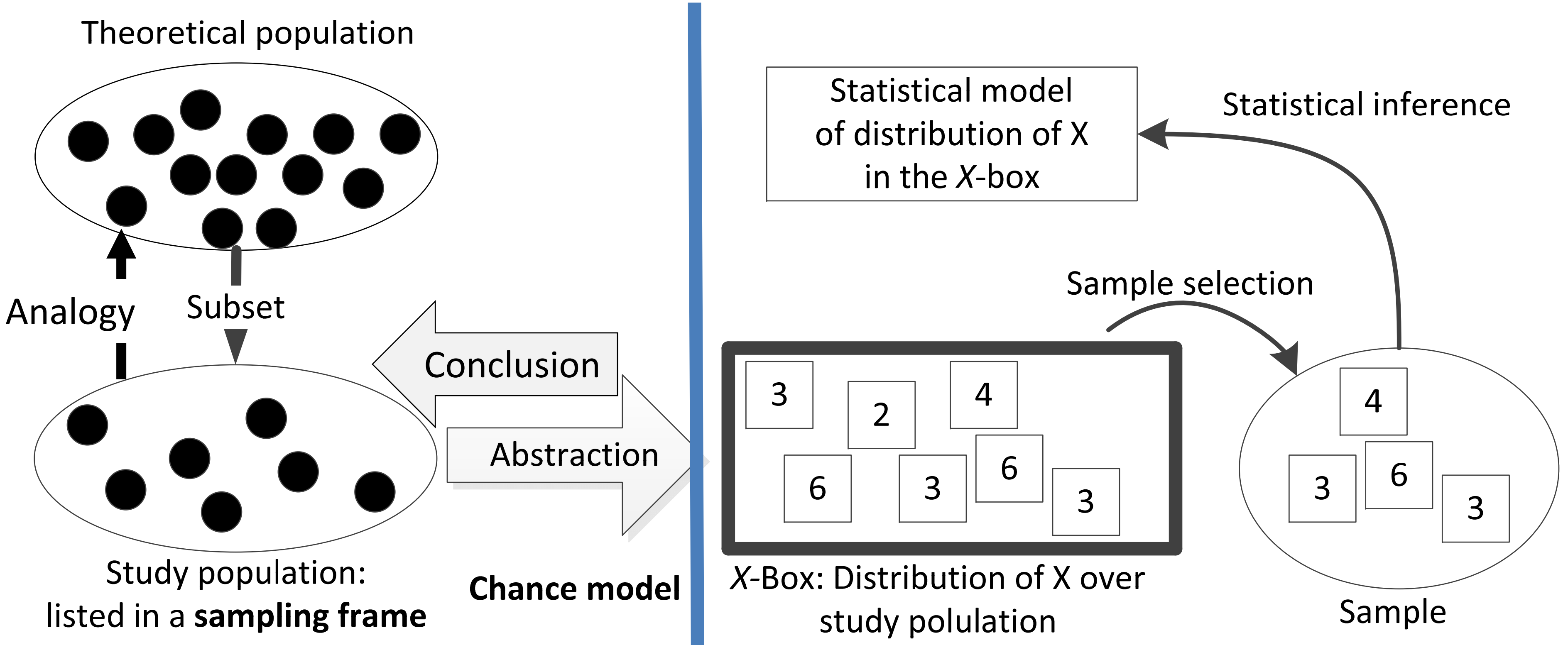
Research methodology.

- Sampling frame,
- Chance model,
- Conclusion

Statistical inference.

- Unobservable distribution of numbers,
- Sample selection,
- Conclusion about unobservable distribution of numbers

Methodology of statistical inference



Research methodology.

- Sampling frame,
- Chance model,
- Conclusion,
- Analogy.

Statistical inference.

- Unobservable distribution of numbers,
- Sample selection,
- Conclusion about unobservable distribution of numbers

Four methods for statistical inference

1. **By big data:** If the sample is almost the size of the population, then the population probably has similar statistics.
   - Only true if the sample is random. Law of large numbers.
2. **By statistical learning:** Use a sample of (X, Y) values to estimate Y as a function of X in the population.
   - E.g. regression. Different methods come with different assumptions.
3. **Bayesian inference.** Use a sample to update a hypothesized population distribution.
   - Need to start with an initial hypothesized distribution.
4. **Frequentist statistical inference:** In repeated random sampling from the same population, the sample averages are approximately normally distributed around the population mean.
   - Central-limit theorem. Assumes random samples.

Four varieties of frequentist statistical inference

- Fisher: Test a null hypothesis
- Neyman-Pearson: Decide between alternative hypotheses
- Neyman: Estimate a confidence interval
- Social sciences: Null Hypothesis Significance Testing (NHST)

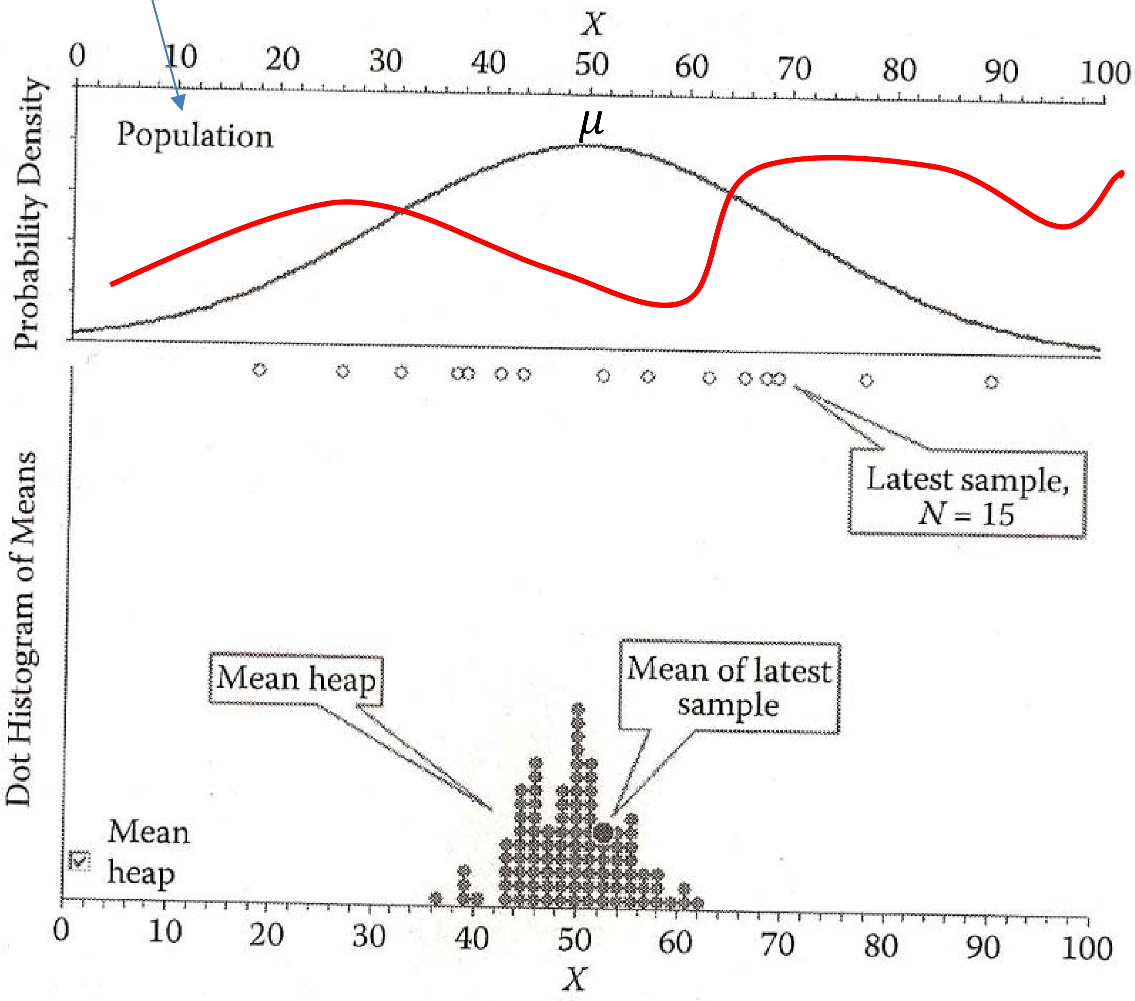
Central-Limit Theorem

- Let $X_1, \dots, X_n$ be a sample from a distribution with mean μ and standard deviation σ. Then the sample mean $\overline{X}_n$ is approximately normally distributed with mean μ and standard deviation${}^{\sigma^2}/_{n}$. The approximation gets better as n gets larger.

Illustration of CLT

!!!!!

He means "distribution of X over the study population"



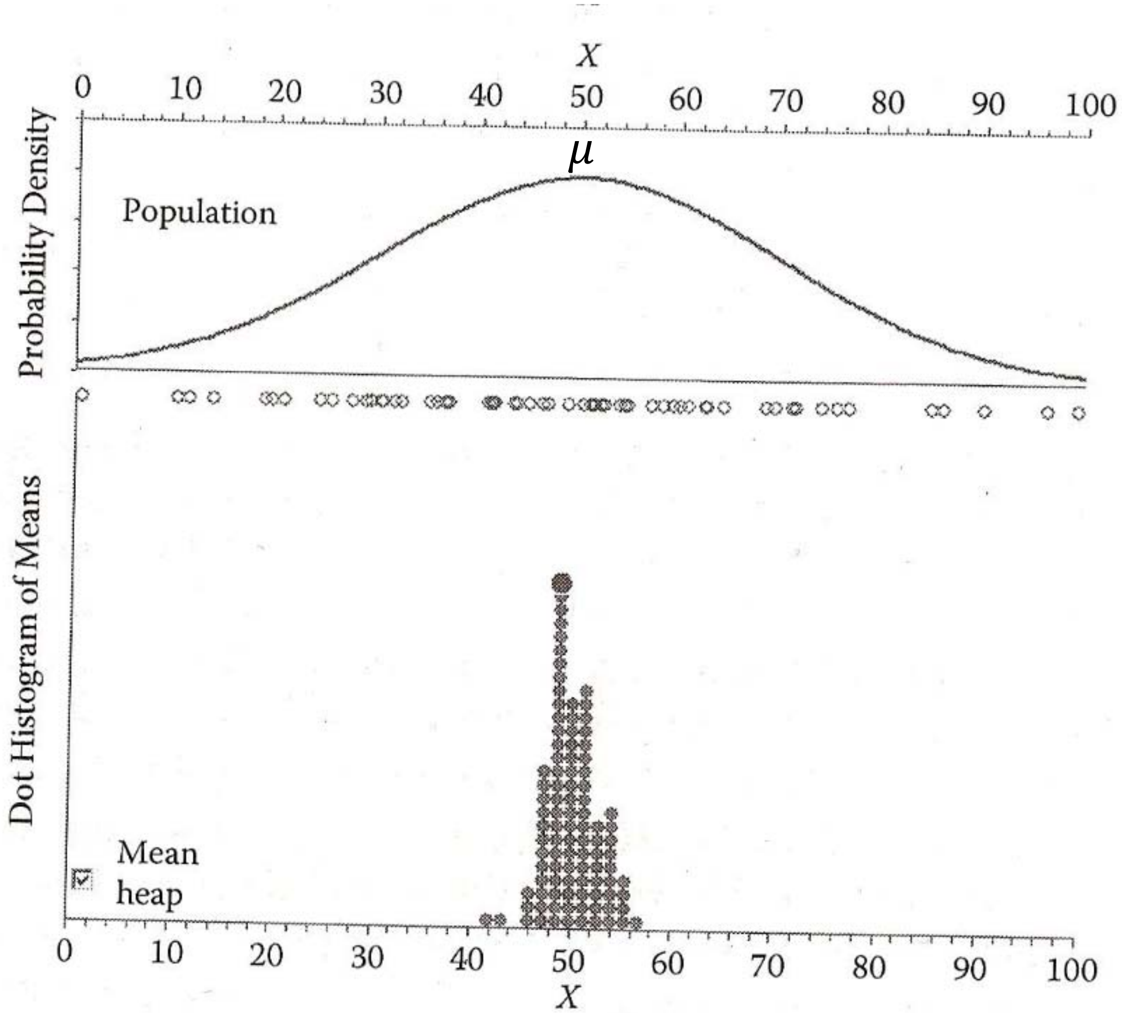
- Distribution does not need to be normal!
- Must have finite σ and μ.

- Repeated random sampling will give an approximately normal distribution of sampling means, centering on the population mean μ, with **standard error** of $\sigma/\sqrt{n}$.

"Error" = Fluctuation due to random sampling

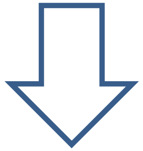
G. Cumming. *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis.* Routledge 2012.

Illustration of CLT



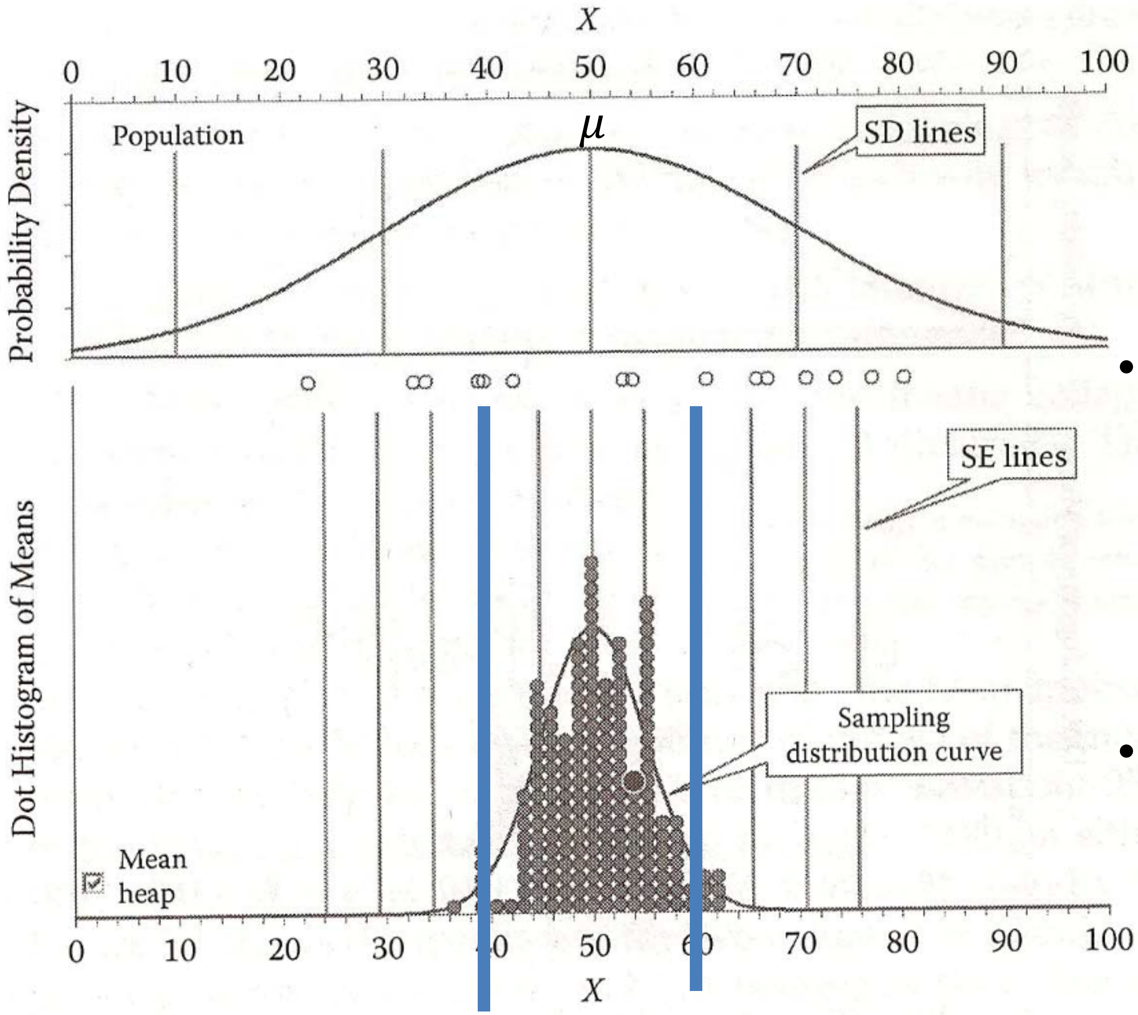
Population distribution does not need to be normal!

Bigger samples (size 60)



Narrower distribution of sampling means

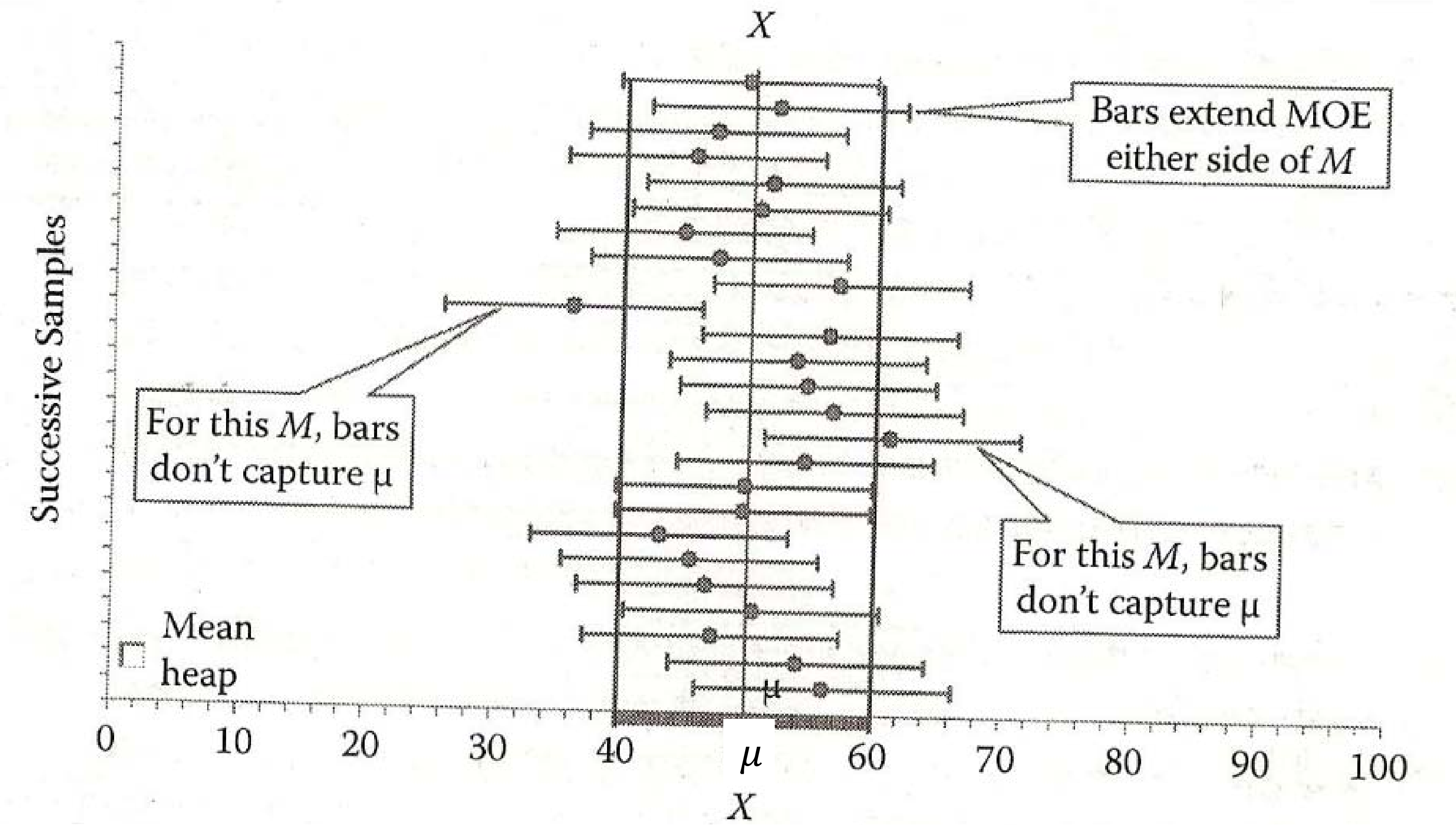
G. Cumming. *Understanding the News Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis.* Routledge 2012.



- Margin of error (MOE): 2 SEs on each side of the population mean
- 5% chance that the sample mean is further away from μ

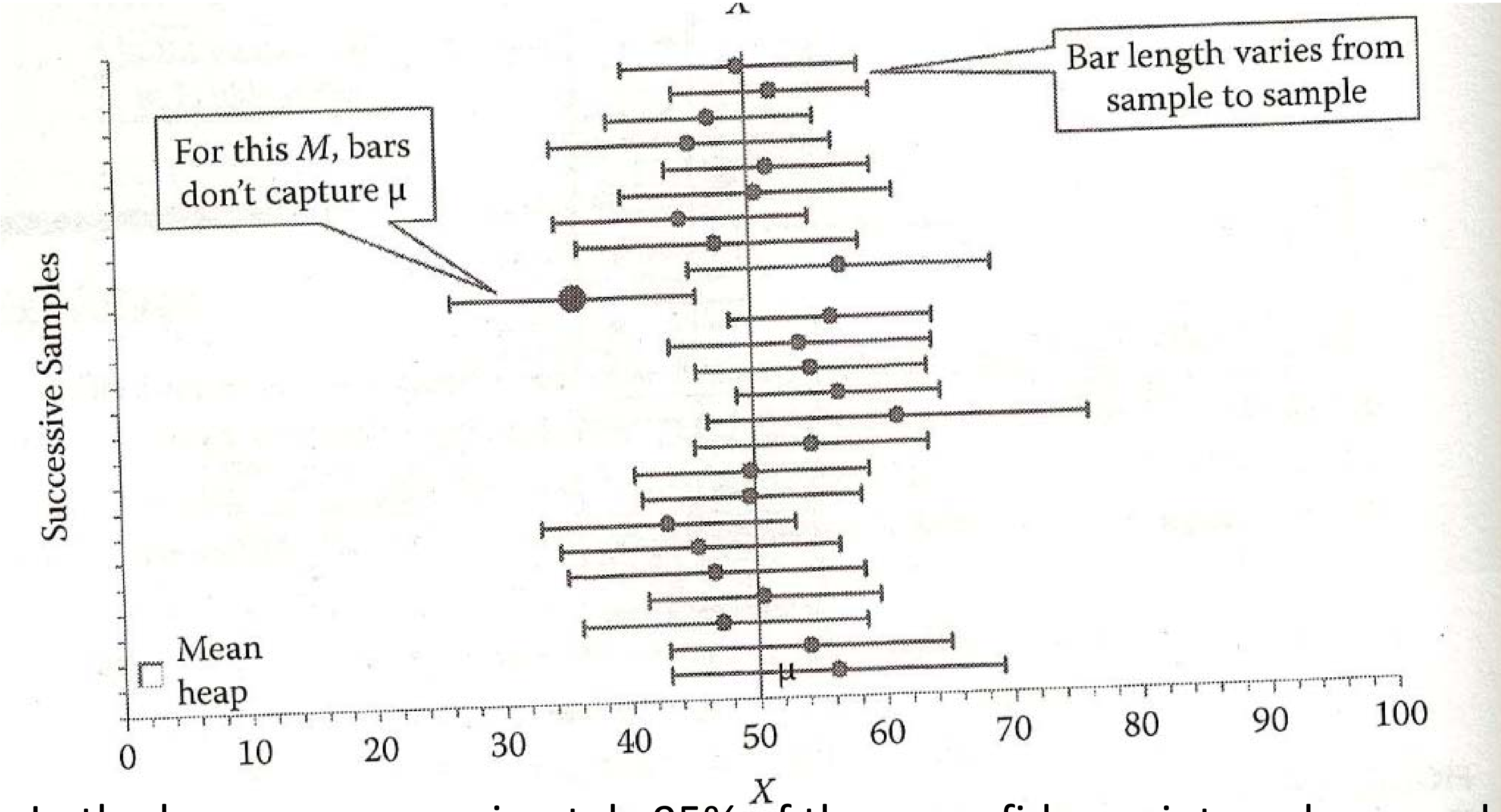
FIGURE 3.4
The upper panel displays the population distribution, with lines marking SD units, showing $\sigma = 20$. Below is the mean heap. The superimposed curve is the sampling distribution of the mean, with lines marking SE units. In this example, $N = 15$, and 200 samples have been taken. The SE $= \sigma/\sqrt{N} = 20/\sqrt{15} = 5.16$.

95% confidence intervals extend 2 SE's around the sample mean



In the long run, 95% of these confidence intervals around the sample mean, will contain the population mean μ

Confidence intervals using $\frac{s}{\sqrt{n}}$ to estimate $\frac{\sigma}{\sqrt{n}}$



In the long run, approximately 95% of these confidence intervals around the sample mean, will contain the population mean μ

95% confidence interval estimation

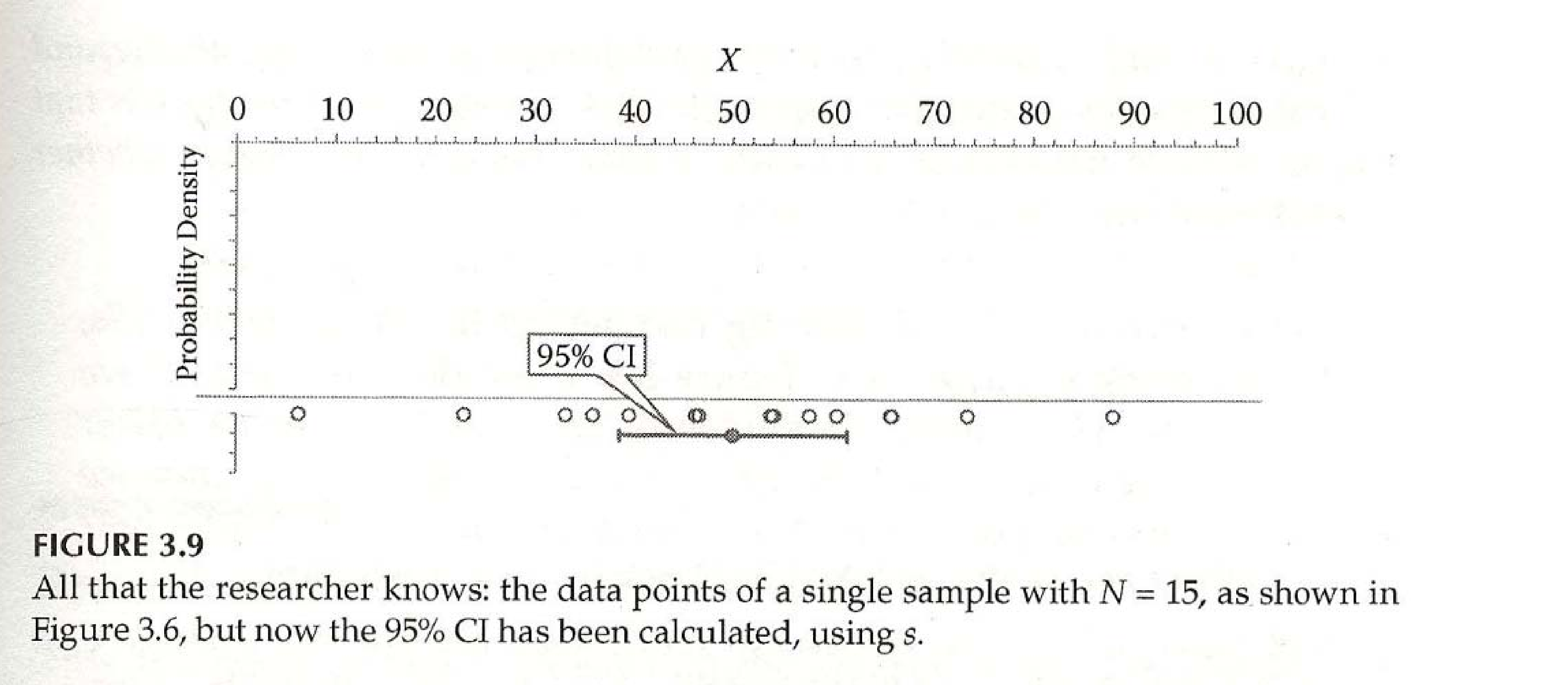


FIGURE 3.9
All that the researcher knows: the data points of a single sample with $N = 15$, as shown in Figure 3.6, but now the 95% CI has been calculated, using s.

In the long run, approximately 95% of these confidence intervals around the sample mean, will contain the population mean μ. We do not know if the current estimation contains μ.

Example: observational study

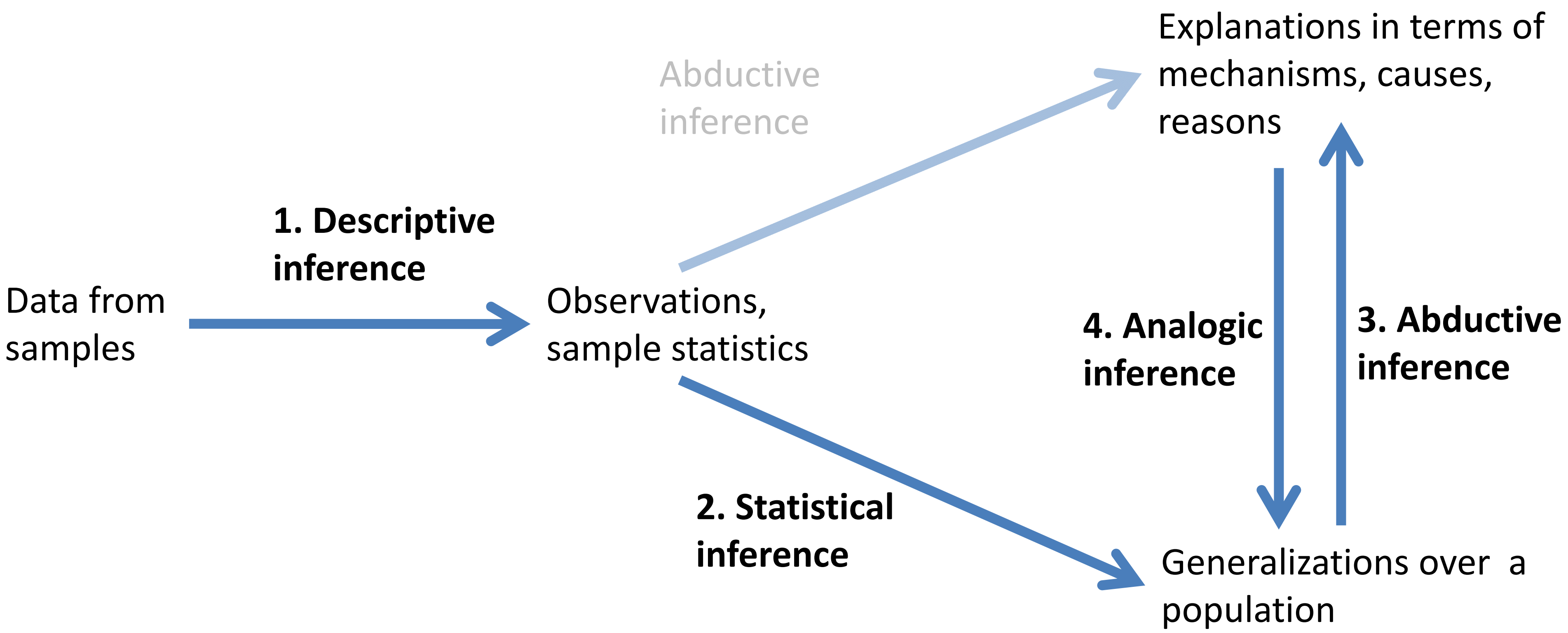
- T. Huynh, J. Miller, An empirical investigation into open source web applications' implementation vulnerabilities. Empir. Softw. Eng. **15**(5), 556–576 (2010)
- *Sample of 20 open source web applications selected randomly from the OS web applications in the OSVDB that satisfies a number of conditions: more than one release, larger than 3 KLOC, exploitable vulnerabilities, available source code.*
  - **(What is the theoretical population?**
  - **What is the study population?)**
- *Count the number of implementation vulnerabilities*
- *Observation: The average percentage of implementation vulnerabilities per OS web application* ***in the sample*** *is 73%.*

- *Statistical inference:*
  - *Assuming a random sample, and*
  - *assuming that the proportion of coding errors is constant and independent across web applications,*
  - *the average percentage of vulnerabilities caused by coding errors in any OS web application* ***in the study population*** *is roughly 73% ± 4% with roughly 95% confidence.*

The paper ends here

- *Abduction (inference to the best explanation):*
  - *Coding errors that cause implementation vulnerabilities are caused by cognitive limitations and project coordination mechanisms .*
  - **(Which ones?)**
- *Analogic generalization to theoretical population:*
  - *The cognitive mechanisms that produce these coding errors and project coordination mechanisms (whatever they are) are common across all web application programmers.*
  - *If there are no interfering mechanisms, then it is plausible that 73% of all vulnerabilities in web applications are implementation vulnerabilities.*

Sample-based inference



Explanations in terms of mechanisms, causes, reasons

Abductive inference

1. **Descriptive inference**

Data from samples

Observations, sample statistics

4. Analogic inference

3. Abductive inference

2. Statistical inference

Generalizations over a population

© R.J. Wieringa

Fisher hypothesis test

Context: scientific reasoning about an unknown distribution mean

1. State a hypothesis about the distribution mean
2. Collect data
3. if your hypothesis is outside the confidence interval for the mean of the data, your data does not support the hypothesis

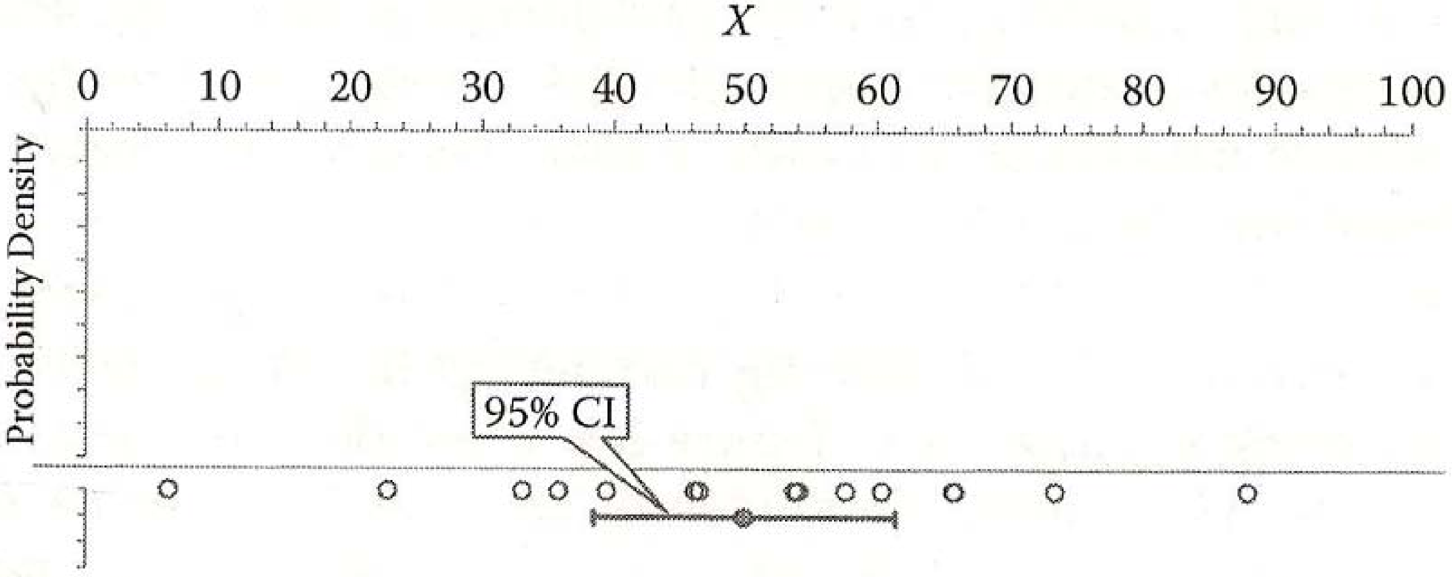


FIGURE 3.9
All that the researcher knows: the data points of a single sample with $N = 15$, as shown in Figure 3.6, but now the 95% CI has been calculated, using s.

Usually, the **p-value** is computed: probability to find the sample mean or a mean further away from the hypothesized distribution mean

Interpretation of the outcome of a Fisher hypothesis test

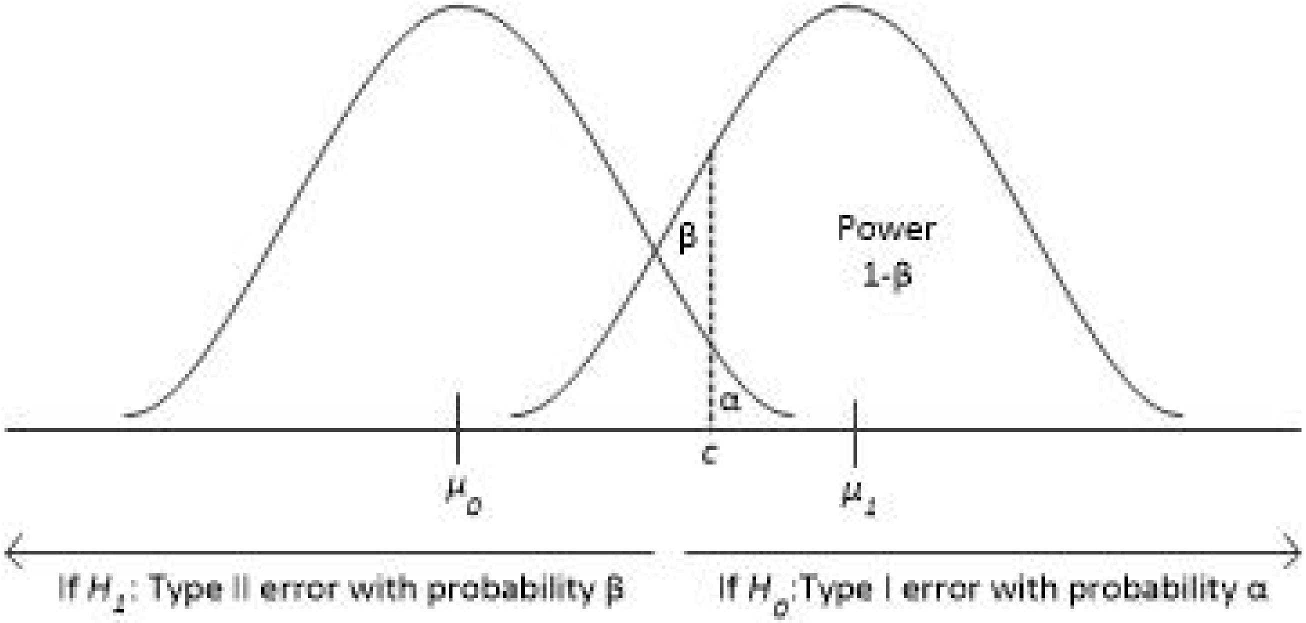
- If your hypothesized μ is inside the confidence interval:
  - The data are consistent with your hypothesis.
- If μ is outside the confidence interval:
  - Perhaps H_0 is false. Your hypothesis about μ is false, so your p-value is wrong and you cannot compute the correct p-value.
  - Perhaps H_0 is true. Then your p-value is correct and we have made a rare observation.
  - Perhaps we have observed an outlier: our data are incorrect.

- Response in all cases: replicate!

Neyman-Pearson hypothesis-testing

- Context: Decide whether H_0 or H_1 is true.
  - Assess the cost of wrong decisions.
  - Set error rates α (probability of incorrectly rejecting H_0) and β (probability of incorrectly rejecting H_1).
  - Set a decision criterion according to these rates.
  - Start deciding this way

In the long run, you will achieve these error rates.


β
Power
1-β
α
c
μ0
μ1
If H1: Type II error with probability β
If H0:Type I error with probability α

Discussion

- Neyman-Pearson inference is appropriate for repeated decisions in which you want to manage your long-run error rates
  - *E.g. quality control,*
  - *Biometrics*
  - **Inductive behavior:** You do not believe anything based on the test outcomes; rather, you start making decisions in a certain way.
- Fisher inference is appropriate for scientific hypothesis testing, which test your beliefs, and where a test may never be repeated.
  - **Null hypothesis testing:** Test whether a hypothesis that you hope is false is incompatible with the data.

Null-hypothesis significance test (NHST): if your H_0 is outside a 95% confidence interval, reject H_0 at 5% level and accept H_1

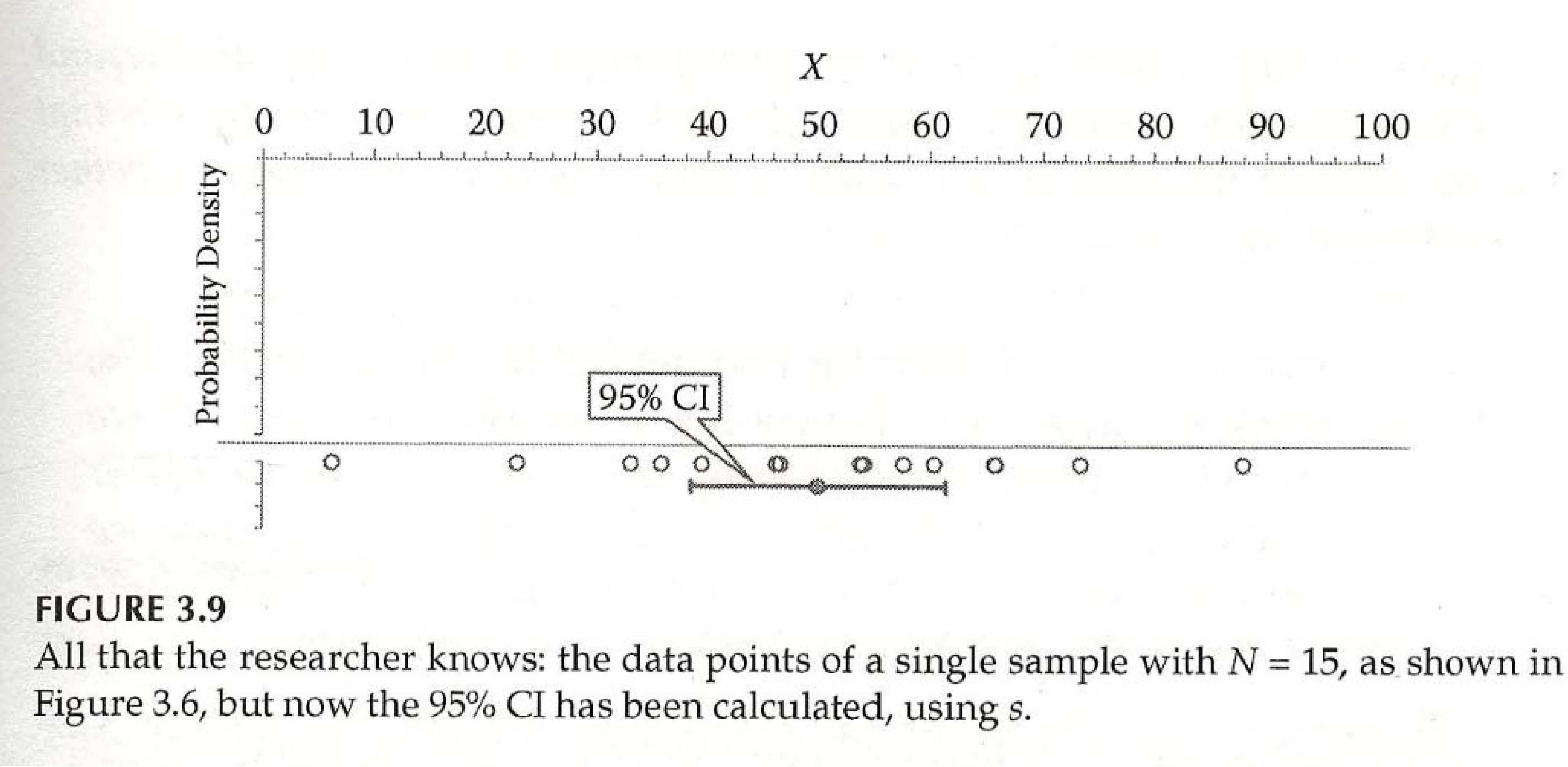


FIGURE 3.9

All that the researcher knows: the data points of a single sample with $N = 15$, as shown in Figure 3.6, but now the 95% CI has been calculated, using *s*.

Misconception 1 of NHST: Fixed decision rule

- Why 5%? What if p-value = 4.9% or 5.1%?
  - Outcome of hypothesis test should be combined with what we know from earlier tests and from established theory.
- Impact of NHST rule:
  - Published p-values crowd just below 5% (“p-hacking’’).
  - Just above 5% they are sparse (‘’publication bias’’)

Misconception 2 of NHST: Probabilistic falsification

- Rule of falsification
  - If $p \rightarrow q$ and we observe $\neg q$, then $\neg p$.
- There is no valid rule of probabilistic falsification
  - If p *probably implies* q and we observe $\neg q$, then no conclusion.

Misconception 3 of NHST: If H_0 is false, then H_1 is true

- There are many alternatives to H_0!
- In NHST, H_0 is **not a substantial hypothesis** but a hypothesis of no difference
  - If we reject H_0 then we can only conclude that “something is going on”
  - But we knew this already.

Causal reasoning using NHST

1. Draw random sample from study population
2. Allocate treatments T_1 and T_2 randomly to sample elements.
3. Apply treatments and measure outcome variable X.
4. Compute p-value of $\mathrm{d} = \bar{X}_1 - \bar{X}_2$ assuming the null hypothesis that $\mathrm{d} = 0$.
5. Statistical inference:
   - If p-value < 5%, the difference is "**statistically significant**", i.e. statistically discernable. Conclude that $\mathrm{d} \neq 0$ in the population.
   - Otherwise conclude that $\mathrm{d} = 0$ in the population.
6. Causal inference:
   - If there are no other, more plausible causal explanations of a difference in outcomes, conclude that it is **caused** by the difference between treatments.

Problems with this use of NHST

- There is always a difference; it would be a miracle it the sample means were identical.
- If the sample is large enough, any difference can be discerned statistically. Follows from CLT.

Example: experimental study

- *Four groups of 9 to 26 students made UML domain model from Use case model for two systems, with or without using System Sequence Diagrams (SSDs) and System operations contracts (SOCs). Four-group crossover design.*
  - *Theoretical population: all software engineers*
  - *Smaller theoretical population: all software engineering students*
  - *Study population: all participants in an SE class*
  - *Sample: Self-selected sample of volunteers*
  - *Groups within this sample: students randomly allocated to UML or to UML+SSD+SOC*

Example continued

- *Observation:*
  - ***In the observed samples**, when SSDs and SOCs were used, average correctness of models was higher, and effort to produce them was lower.*
- *Generalization by NHST:*
  - *Pairwise t-test, simple repeated measures ANOVA and mixed repeated measures ANOVA support the generalization that average correctness of models and effort to produce them is better when SSDs and SOCs are used **in the population** of all software engineering students. This conclusion is plausible but not always correct.*
- *Explanation:*
  - *By listing all possible causes, and assessing them on their plausibility, the use of SSDs and SOCs is the most plausible cause of these effects (and not the competence of the students or the positive expectation of the experiments, or …)*
- *Generalization by analogy to similar populations, e.g. the population of all SE students or of professional software engineers.*
  - *Need to discuss if the social or cognitive mechanisms that produce the results in the student population, are the same as those in the theoretical population of all SE students or of all professional software engineers.*

An aside

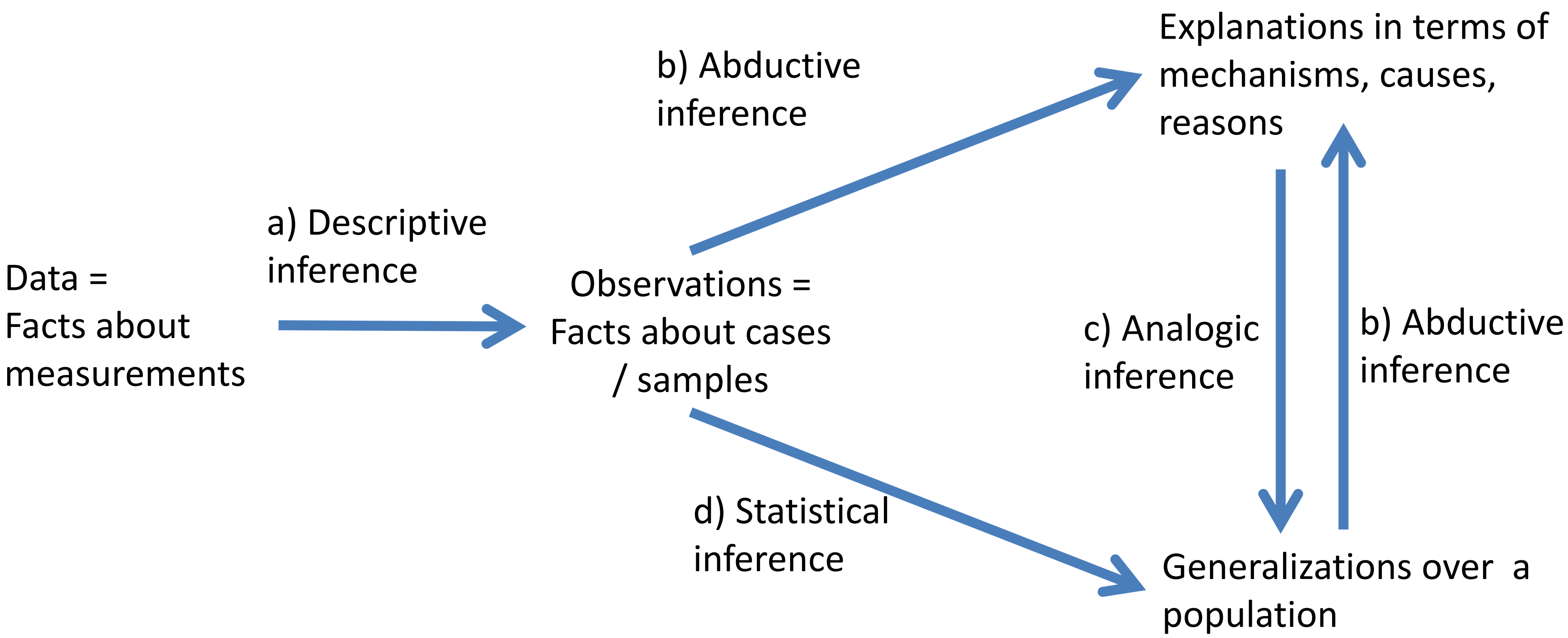
- L. Briand, Y. Labiche, R, Mardazo-Rivera. “An experimental evaluation of the impact of systems sequence diagrams and system operation contracts on the quality of the domain model”. *ESEM 2011,* Page 157-166. ACM Press.
- They did this ….. but unfortunately found hardly any support for a statistically significant difference.

Statistical conclusion validity

- **Statistical conclusion validity** = Degree of support for a statistical inference

- **Stable distribution.** Does X have a stable distribution, with fixed parameters?
- **Sampling.** Is sample selection random?
- **Treatment allocation.** Are treatments allocated randomly to sample elements?
- **Scale.** Does X have an interval or ratio scale?
- Assumptions of particular statistical techniques

Validity of inferences



Data =
Facts about
measurements

a) Descriptive
inference

Observations =
Facts about cases
/ samples

b) Abductive
inference

Explanations in terms of
mechanisms, causes,
reasons

c) Analogic
inference

b) Abductive
inference

d) Statistical
inference

Generalizations over a
population

a) **Descriptive validity**: no information added in the descriptions
b) **Internal validity**: degree of support for explanations
c) **External validity**: degree of support for analogic generalizations
d) **Statistical conclusion validity**: degree of support for statistical inference

Outline

- **Introduction**
  - The design cycle
  - Theories
  - The research setup
- **Scientific inference**
  - Description
  - Explanation
  - Generalization
  - Prediction
- **Empirical research**
  - Checklist
  - Example research methods

- Predictions predict effects of events, causes, mechanisms or reasons.
  - **Empirical prediction:** use a case description or statistical property of a population to predict future events. Must be based on analogy or statistical inference.
  - **Causal prediction:** Use cause-effect relationship to predict the effect of a treatment. Must be based on experimental evidence.
  - **Architectural prediction:** Use mechanism to predict the effect of a stimulus on a case. Must be based on architectural analysis.
  - **Rational prediction:** Use goals and motivations to predict what an actor will do. Must be based on assumptions about rationality, goals, motivation.

Explanations versus predictions

- Explanations are about the past, predictions are about the future
- Explanations may not allow prediction, because they may require knowledge we do not have in advance of the predicted event.
  - *Explanations of the outcome of a football match*

- Predictions may be unexplainable, because they are based on observed regularities, without sufficient understanding of mechanisms.
  - *Weather forecast*
- Predictions can be empirically tested.
  - Repeated experiments can provide solid evidence for a prediction
  - Explanations may change, validated predictions do not.

Usable predictions

- Designers produce theories of the form Artifact x Context →Effect.
- A practitioners can use this in their context if
  - They can acquire the Artifact (budget, time)
  - They can recognize that their case matches Context

    Usable
  - They want to achieve Effect (goals, law, ethics)
  - There are no additional, unwanted effects of A x C.

    Useful

Outline

- **Introduction**
  - The design cycle
  - Theories
  - The research setup
- **Scientific inference**
  - Description
  - Explanation
  - Generalization
  - Prediction
- **Empirical research**
  - Checklist
  - Example research methods

A classification of research setups

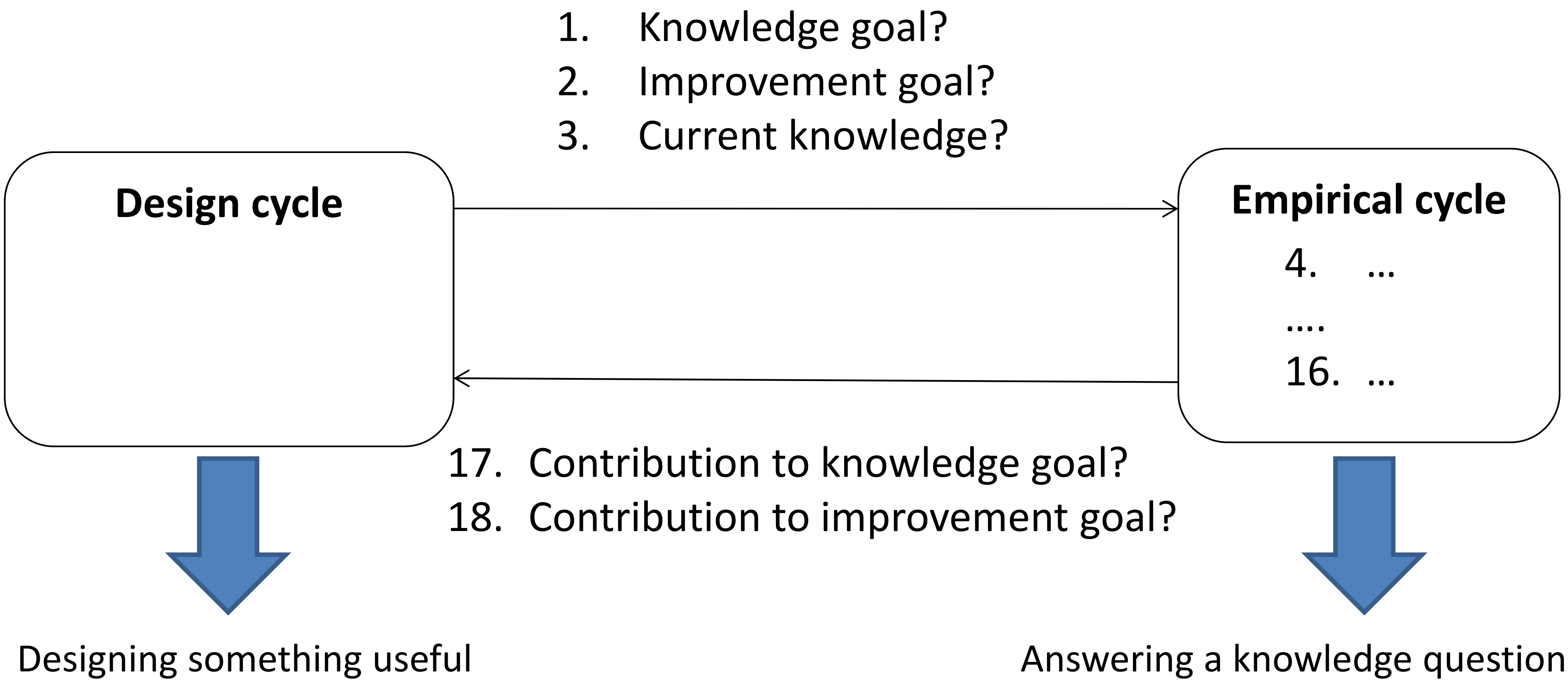
	Observational study (no treatment)	**Experimental study (treatment)**
Case-based: investigate single cases, look at architecture and mechanisms.	**Observational case study**	• **Expert opinion** (mental simulation by experts), • **Mechanism experiment** (simulation, prototyping), • **Technical action research** (experimental use of the artifact in the real world)
Sample-based: investigate samples drawn from a population, look at averages and variation	**Survey**	• **Statistical difference-making experiment** (treatment group – control group experiments)

Third dimension: lab or field

Different designs support different inferences

	Observational study (no treatment)	Experimental study (treatment)
Case-based: investigate single cases, look at architecture and mechanisms.	**Observational case study**	• **Expert opinion** (mental simulation by experts), • **Mechanism experiment** (simulation, prototyping), • **Technical action research** (experimental use of the artifact in the real world)
	Evidence for or against architectural theories of similar cases	
Sample-based: investigate samples drawn from a population, look at averages and variation	**Survey:** *Evidence for or against estimations of properties of population distributions*	• **Statistical difference-making experiment** (treatment group – control group experiments): *Evidence for or against causal theories*

Checklist for the empirical cycle: context



1. Knowledge goal?
2. Improvement goal?
3. Current knowledge?

Design cycle

Empirical cycle

4. …

….

16. …

17. Contribution to knowledge goal?
18. Contribution to improvement goal?

Designing something useful

Answering a knowledge question

- Checklist for design, reporting, reading.

This is a checklist for

- research design,
- research reporting,
- reading a report.

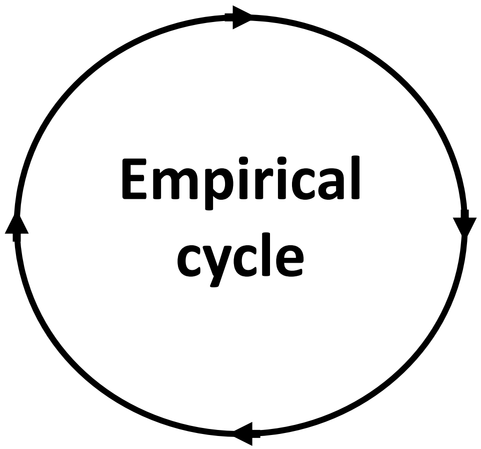
App. B in my book & my web site

Data analysis

12. Descriptions?
13. Statistical conclusions?
14. Explanations?
15. Generalizations?
16. Answers?

Research execution

11. What happened?



Empirical cycle

Research problem analysis

4. Conceptual framework?
5. Knowledge questions?
6. Population?

Design validation	**Research & inference design**	
7. Object of study validity?	7. Object of study?	Research setup
8. Treatment specification validity?	8. Treatment specification?	
9. Measurement specification validity?	9. Measurement specification?	
10. Inference validity?	10. Inference?	Inference

Comparison with other checklists

- A. Jedlitschka and D. Pfahl, “Reporting guidelines for controlled experiments in software engineering,” in *Proceedings of the 4th International Symposium on Empirical Software Engineering (ISESE 2005)*. IEEE Computer Society, 2005, pp. 94–104.
- P. Runeson and M. Höst, “Guidelines for conducting and reporting case study research in software engineering,” *Empirical Software Engineering*, vol. 14, pp. 131–164, 2009.
- K. Schulz, D. Altman, and D. M. D, “CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials,” *Annals of Internal Medicine*, vol. 152, no. 11, pp. 1–7, 1 June 2010.
- **Comparison in** Wieringa, R.J. (2012) *A Unified Checklist for Observational and Experimental Research in Software Engineering (Version 1).* Technical Report TR-CTIT-12-07, Centre for Telematics and Information Technology, University of Twente, Enschede. http://eprints.eemcs.utwente.nl/21630/

Outline

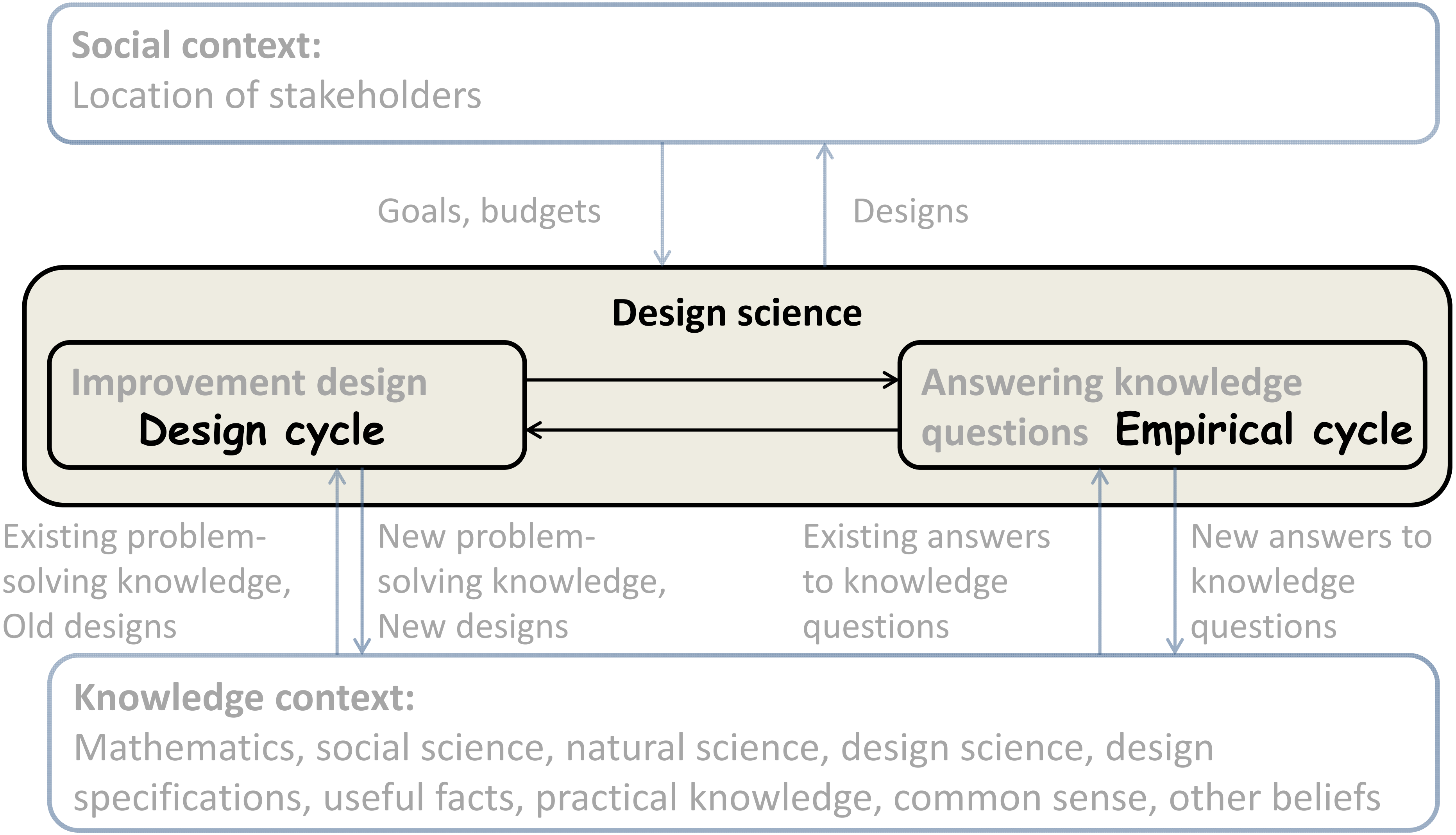
- **Introduction**
  - The design cycle
  - Theories
  - The research setup
- **Scientific inference**
  - Description
  - Explanation
  - Generalization
  - Prediction
- **Empirical research**
  - Checklist
  - Example research methods

Example case-based research methods

Separate slide set.

- Observational case study
- Single-case experiment
- Multiple-case experiment
- Technical action research

Framework for design science



Social context:
Location of stakeholders

Goals, budgets

Designs

Design science

Improvement design
Design cycle

Answering knowledge questions
Empirical cycle

Existing problem-solving knowledge, Old designs

New problem-solving knowledge, New designs

Existing answers to knowledge questions

New answers to knowledge questions

Knowledge context:
Mathematics, social science, natural science, design science, design specifications, useful facts, practical knowledge, common sense, other beliefs

© R.J. Wieringa

Take-home

- Design theories are about the effects of an artifact in a context
- Theory consists of conceptual framework and generalizations
- Explanations can be causal, architectural, rational
- Generalization can be case-based (analogic) or sample-based (statistical)
- Theories are fallible and must be assessed on validity

© R.J. Wieringa

- Wieringa, R.J. and Daneva, M. (2015) *Six strategies for generalizing software engineering theories.* Science of computer programming, 101. pp. 136-152.

- Wieringa, R.J. (2014) *Design science methodology for information systems and software engineering.* Springer Verlag

- Wieringa, R.J. (2014) *Empirical research methods for technology validation: Scaling up to practice.* Journal of systems and software, 95. pp. 19-31.

- Wieringa, R.J. and Morali, A. (2012) *Technical Action Research as a Validation Method in Information Systems Design Science.* In: *Design Science Research in Information Systems. Advances in Theory and Practice 7th International Conference*, DESRIST 2012, 14-15 May 2012, Las Vegas, USA. pp. 220-238. Lecture Notes in Computer Science 7286. Springer.

- Wieringa, R.J. (2010) *Relevance and problem choice in design science.* In: *Global Perspectives on Design Science Research (DESRIST). 5th International Conference*, 4-5 June, 2010, St. Gallen. pp. 61-76. Lecture Notes in Computer Science 6105. Springer.

- Wieringa, R.J. (2009) *Design Science as Nested Problem Solving.* In*: Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology*, Philadelphia. pp. 1-12. ACM.