[ simula . research laboratory ]

# Research Methods I

## a) Introduction to scientific method
## b) Experiments
## c) Surveys

**Learning goals**: Improved ability to understand, design and evaluate research studies based on experimental methods.

**Supporting texts**:

• Briony J Oates. Researching Information Systems and Computing (Section 3: Overview of the research process, Section 7: Surveys, and Section 9: Experiments)

• Barbara Kitchenham et al., Preliminary Guidelines for Empirical Research in Software Engineering, IEEE Transactions on Software Engineering, 2002.

• www.freeinquiry.com/intro-to-sci.html

# Science - Wikipedia

**Science** (from the Latin scientia, 'knowledge'), in the broadest sense, refers to any systematic knowledge or practice.[1] In a more restricted sense, science refers to a system of acquiring knowledge based on the scientific method, as well as to the organized body of knowledge gained through such research. *[text deleted]*

Fields of science are commonly classified along two major lines:

Natural sciences, which study natural phenomena (including biological life), and

Social sciences, which study human behavior and societies.

These groupings are empirical sciences, which means the knowledge must be based on observable phenomena and capable of being experimented for its validity by other researchers working under the same conditions.[4]

Mathematics, which is sometimes classified within a third group of science called formal science, has both similarities and differences with the natural and social sciences. It is similar to empirical sciences in that it involves an objective, careful and systematic study of an area of knowledge; it is different because of its method of verifying its knowledge, using a priori rather than empirical methods. *[text deleted]* The formal sciences are essential in the formation of hypotheses, theories, and laws,[6] both in discovering and describing how things work (natural sciences) and how people think and act (social sciences).

Science *[text deleted]* is sometimes termed experimental science to differentiate it from applied science, which is the application of scientific research to specific human needs, though the two are often interconnected.

[ simula . research laboratory ]

# What is science?

**Important elements of science (most researchers will agree on these):**

- Empirical evidence (exception for mathematics?)
- Logical reasoning
- Skeptical attitude

The following slides describes professor **Steven D. Schafersman's** viewpoints on these elements. He is a geologist**,** i.e., is from "natural sciences".

Many researchers will not agree with him in everything he claims. His viewpoints, however, are typical for scientists with a strong "positivistic" (more on this later) attitude and represent well the "traditional" view on science.



# What is science?
## www.freeinquiry.com/intro-to-sci.html

**The Use of Empirical Evidence**

- *"Empirical evidence is evidence that one can see, hear, touch, taste, or smell;* it is evidence that is susceptible to one's senses. Empirical evidence is important because it is evidence that others besides yourself can experience, and it is repeatable, so empirical evidence can be checked by yourself and others after knowledge claims are made by an individual. Empirical evidence is the *only* type of evidence that possesses these attributes and is therefore the only type used by scientists and critical thinkers to make vital decisions and reach sound conclusions."

## What is science?
## www.freeinquiry.com/intro-to-sci.html

**Rationalism: The Practice of Logical Reasoning**

- "Scientists and critical thinkers always use logical reasoning. *Logic allows us to reason correctly*, but it is a complex topic and not easily learned; many books are devoted to explaining how to reason correctly, and we can not go into the details here. However, I must point out that most individuals do not reason logically, because they have never learned how to do so. Logic is not an ability that humans are born with or one that will gradually develop and improve on its own, but is a skill or discipline that must be learned within a formal educational environment. Emotional thinking, hopeful thinking, and wishful thinking are much more common than logical thinking, because they are far easier and more congenial to human nature. Most individuals would rather believe something is true because they feel it is true, hope it is true, or wish it were true, rather than deny their emotions and accept that their beliefs are false.2



## What is science?
## www.freeinquiry.com/intro-to-sci.html

**Skepticism: Possessing a Skeptical Attitude**

- "The final key idea in science and critical thinking is skepticism, the *constant questioning of your beliefs and conclusions*. Good scientists and critical thinkers constantly examine the evidence, arguments, and reasons for their beliefs. Self-deception and deception of yourself by others are two of the most common human failings. Self-deception often goes unrecognized because most people deceive themselves. The only way to escape both deception by others and the far more common trait of self-deception is to repeatedly and rigorously examine your basis for holding your beliefs. You must question the truth and reliability of both the knowledge claims of others and the knowledge you already possess. One way to do this is to test your beliefs against objective reality by predicting the consequences or logical outcomes of your beliefs and the actions that follow from your beliefs. If the logical consequences of your beliefs match objective reality--as measured by empirical evidence--you can conclude that your beliefs are reliable knowledge (that is, your beliefs have a high probability of being true)."

## Why do science?
## www.freeinquiry.com/intro-to-sci.html

"Science has unquestionably been the most successful human endeavor in the history of civilization, because it is the only method that successfully discovers and formulates reliable knowledge.

The evidence for this statement is so overwhelming that many individuals overlook exactly how modern civilization came to be (our modern civilization is based, from top to bottom, on the discoveries of science and their application, known as technology, to human purposes.).

Philosophies that claim to possess absolute or ultimate truth invariably find that they have to justify their beliefs by faith in dogma, authority, revelation, or philosophical speculation, since it is impossible to use finite human logic or natural evidence to demonstrate the existence of the absolute or ultimate in either the natural or supernatural worlds.

Scientific and critical thinking require that one reject blind faith, authority, revelation, and subjective human feelings as a basis for reliable belief and knowledge. These human cognitive methods have their place in human life, but not as the foundation for reliable knowledge."



## Research paradigms
## (based on the Briony Oates' text-book)

- Positivism
  - Controlled experiments, surveys, case studies, action research
- Interpretive research
  - Ethnography, case studies, action research, surveys
- Critical research
  - Action research, ethnography, case studies

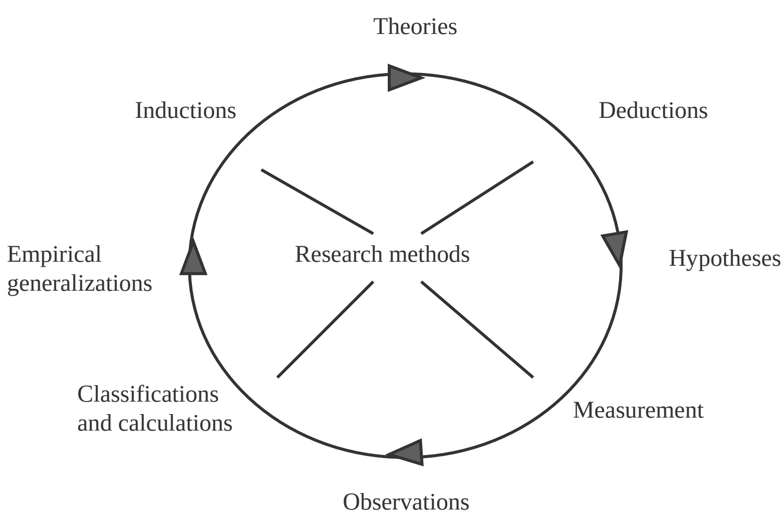
**NB**: The above paradigms focus on theory building and testing. In addition, we may add "constructive research". This type of research includes many (most?) software engineering research papers and aims at constructing products or methods scientifically. This type of research is not the focus of our lectures.

## Positivism

- Originally developed for the use in natural science, i.e., not studies of human behavior.
- Knowledge generation through Wallace's cycle (see next slide).
- Based on reductionism, repeatability and refutation (falsification, ref. Popper).
- Assumptions:
  - Our world is ordered, not random
  - We can investigate the world objectively (Well, at least achieve an acceptable degree of "inter-subjectivity".)
- Goal: Discover patterns.
- Criteria:
  - Objectivity (or at least inter-subjectivity)
  - Reliability
  - Internal validity (= the extent to which a study evaluates the intended hypotheses, i.e., that it is not likely that rival hypotheses explains the findings)
  - External validity (= the extent to which the results of a study extend beyond the limited sample used in the study)



## Classical Research Process (Wallace's model)



# Interpretive Research

- "*Interpretive research in IS and computing is concerned with understanding the social context of an information system: the social processes by which it is developed and construed by people and through witch it influences, and is influenced by, it social setting.*" (p 292, in Briony J. Oates)
- Try to identify, explore and explain ("rich understanding") how factors in a particular social setting are related and interdependent. Case studies are typically preferred.
- Characteristics:
  - Multiple subjective realities
  - Dynamic, socially constructed meaning
  - Researcher reflexivity (researchers should reflect on their own assumptions, beliefs and actions and their impact on the research process)
  - Study of people in their natural social setting (typically, case studies)
  - Qualitative data analysis
  - Multiple interpretations

[ simula . research laboratory ]

# Interpretive Research

- Criteria (somewhat forced into a positivistic framework):
  - Trustworthiness (more general concept than validity?)
  - Confirmability (analogue to objectivity - can we follow the arguments from the raw data to the interpretation?)
  - Dependability (analogue to reliability and repeatability – is the research process well documented?)
  - Credibility (analogue to internal validity – is it valid to draw the conclusions based on the data collected?)
  - Transferability (analogue to external validity – is it possible to transfer the findings to other cases?
    - **NB**: This is frequently not a goal in interpretive research. An interesting case is an interesting case, even when not transferable to other cases.

[ simula . research laboratory ]

# Critical Research

- "*Critical research in IS and computing is concerned with identifying power relations, conflicts and contradictions, and empowering people to eliminate them as sources of alienation and domination.*" (p. 296, in Briony J. Oates)
- Characteristics:
  - Emancipation (The goal is not only to understand, but free people from being dominated etc.)
  - Critique of tradition (It is essential to question status quo)
  - Non-performative intent (Critical to research with a focus on managers' need for control and profit)
  - Critique of technological determination (People should be in control of technology development)
  - Reflexivity (Strong focus on own beliefs and values as researcher)



# Which research method and paradigm is best?

- Wrong question! Most research methods and paradigms have their strengths and weaknesses.
- It is the relation between the research method, paradigm and research question (goal of a study) that matters.
- In practice, however, the choice of research method and paradigm is very much determined by personal preference and/or set of personal values (ideals).
  - This has the consequence that the choice of research method may be value based instead of selection of the best suited research methods.
    - Researchers belonging to "interpretive research" may not like to use statistics on people, which is essential among positivists.
    - Researchers belonging to "positivism" may not like the lack of pre-made analysis structure typical for interpretive research.

## (Controlled) Experiment

- Belongs to the positivistic tradition.
- Manipulation of at least one variable, i.e., the “treatment”.
  - Example: Treatment A = Use of XP, Treatment B = Use of the Waterfall model
- Testing of hypotheses.
  - Productivity of XP is higher than productivity of Waterfall model.
  - Independent variable = Development method (XP or Waterfall)
  - Dependent variable = Productivity (“depends” on the development method)
- Strong on cause-effect relationships (mainly when treatment is randomized)
  - Without randomized treatment we have quasi-experiments where we have to argue that there are no alternative explanations.
  - Example: The developers are not randomly assigned to the use of XP or the Waterfall model. Perhaps are those using XP more motivated or more competent?



## Experiment

- Typical process:
  - Hypothesis generation (e.g., derived from theory).
    - For example: Treatment A leads to higher X than treatment B.
  - Design a study where the hypothesis can be tested.
    - Study may, for example, be designed to demonstrate the existence of an effect of treatment, to examine effect size of treatment in realistic settings, or to test the robustness/generality of the effect of an treatment.
    - Study may be conducted in a particular context, have certain task and certain participants. These may be representative, extreme, randomly selected, or, selected by convenience.
  - Allocation of treatment to participant
    - Randomly (eases the cause-effect analysis), self-selected, ...
  - Execution of study, measurement and collection of data
  - Statistical analysis of data.
    - For example: Is the difference in effect statistically significant?
  - Interpretation of results should be done in light of previous results!

[ simula . research laboratory ]

## Evaluation of experiments

- Internal validity (Are there alternative explanations that can explain the results?)
  - Events other than the treatment that could have impacted the outcome?
  - Fatigue confounded the effect of the treatment?
  - Hawthorne effect occurred?
  - Measurement problems?
  - Statistical regression?
  - Biased selection of subjects, or biased allocation of subjects to treatment
  - Different loss of participants in different treatment groups



## Internal validity – Exercises

- In an experiment, the effect of rewards on students' academic test results was evaluated. The hypothesis was that if the students were rewarded for good performance they would be more motivated and perform better on the next tests. (Many parents do exactly this to their children.)
- The experiment was designed as follows:
  1. Completion of Test A by 100 students. They knew that the best half would get rewarded.
  2. The 50 best students were rewarded (given $100 each) for their good performance
  3. Then, completion of Test B by the same 100 students. They also knew that the best half would get rewarded.
- Results:
  - The best students on Test A reduced, on average, their performance on Test B. The worst half improved their performance on Test B compared to Test A.
- Conclusion:
  - Rewards of this type does more harm than good for students' performance.
- **Question**: Are there problems with the internal validity of this study?

## Evaluation of experiments

- External validity:
  - Are the samples representative for the population of interest (the one we want to generalize to)?
    - Participants (When are students representative for software professionals?)
    - Tasks (What can we say about real world tasks based on results from smaller tasks?)
    - Contexts (What can we say about real-world effects from effects in laboratory settings?



## External validity - Exercise

- Design a study were two teaching techniques for learning OO-programming are compared.
  - A: Start early with the concept of "classes"
  - B: Learn simpler concepts, like if-then, while, .... first. Then, learn the OO-stuff.
- Formulate a sufficiently precise research question.
- Design an experiment with an acceptable level of both internal and external validity.

## Survey

- Data collection typically through questionnaires and/or interviews.
- Typically sampling from of a well-defined population.
- If not extensive (all IT-developers in Norway), the survey should include a sampling technique that enables generalization to the population (random sampling, stratified sampling, etc.)
- Frequently, a cheap an simple way to get information. Many MSc students applies this in their master thesis work.
- Easy to conduct, very difficult to conduct high quality surveys!!!!
  - Few surveys use proper sampling techniques



## Evaluation of surveys

To be able to generalize (with confidence) the survey should:

- Be based on a proper sampling technique
- Have a high response rate (or know how to adjust for this):
  - Are the non-responders different from those who responds?
- Ensure that questions are interpreted similarly by all respondents.
  - Is, for example, the term "agile" interpreted similarly?
- Ensure that misunderstands are avoided and that the respondents have the necessary competence and motivation for quality responses.

## Survey Exercise

**Purpose**: Study of change in estimation practice from 5 years ago until today.

**Sample:** Participants at Java Zone 2007 visiting a particular company. Participants were stimulated to participate through the possibility to win a trip to Dublin.

**What is/was typically the main input to the estimating activity?**

Today? [ ] User stories [ ] Use cases [ ] Screen/service specification [ ] Other

5 years ago? [ ] User stories [ ] Use cases [ ] Screen/service specification [ ] Other

**Question**: What are strengths and weaknesses of these questions?