cphbusiness

Experimental approach

Computer science

3rd semester

Autumn 2015

Agenda

cphbusiness

- Student presentations - B
  - Beslutningsteori: Martin Karlsen
  - Valg af metode: Simon T
  - Prototyping: Ib

- Experiments
  - What is an experimental approach?
  - Types of experiments
  - Experiment iterations
  - Complications

- Experiment sprint

Agenda

cphbusiness

- Student presentations – A
  - Ebbe – XP
  - Pelle – Agil
- Experiments
  - What is an experimental approach?
  - Types of experiments
  - Experiment iterations
  - Complications

- Experiments

What is an experimental approach?

cphbusiness

- What is an **experiment**?
- Actions where
  - Knowledge/learning is the reason to perform the action
  - All other results are considered irrelevant

- What is an **experimental approach**?
  - An approach that uses experiments as a central way to obtain knowledge

cphbusiness

Situations where experiments are relevant

- Complexity
  - To much information or a to huge problem
- Uncertainty
  - Lack of or uncertain knowledge
- Pragmatism
  - Accept non-perfect solutions
    - Economic reasons

- Examples you know from PT lectures (Floyd)
  - User requirement - Horizontal prototypes
  - Technological possibilities - Verticals prototypes

cphbusiness

Example

- Technological possibilities

- Complexity
  - We ‘never’ know a technology completely
- Uncertainty
  - We know that parts of the technology are not described and we know that the descriptions have errors
- Pragmatism
  - We know that we ‘can not’ make a - technology -perfect solution

When to use experiments

cphbusiness

- Experiments are the **only source** to knowledge

OR

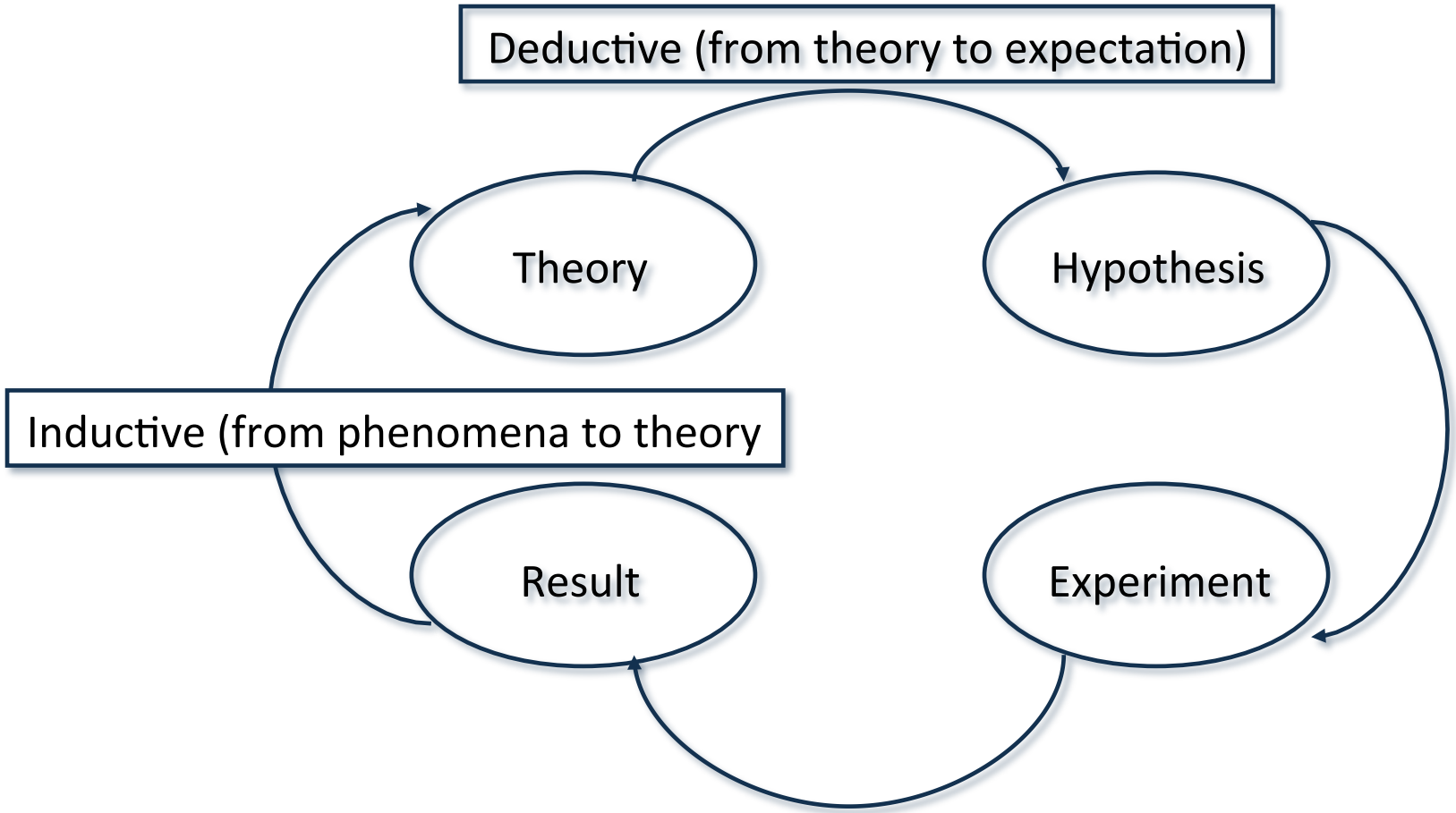
- Experiments are expected to be the **cheapest/fastest** way to get knowledge

OR

- You personally find experiments is a **preferable** way to get knowledge

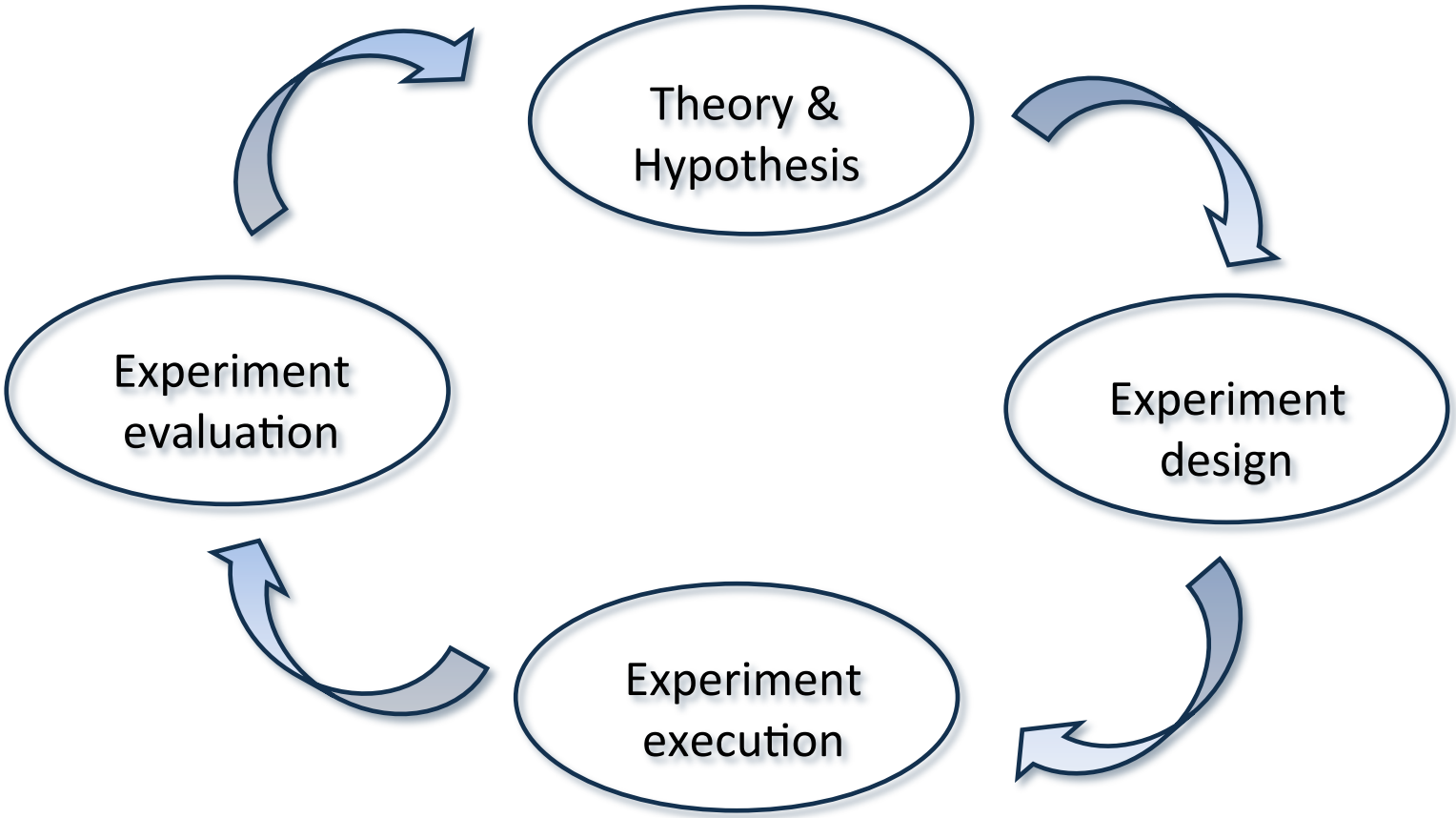
cphbusiness

Experiments in Natural science



cphbusiness

Experiments in systems development



Types of experiments

cphbusiness

- Verification/falsification experiments
  - theses disproved or 'proofed'
  - Remember: To disprove is much easier than proof
- Quantitative experiments
  - Measure the system
  - Remember: Easy to find typical measure but hard to find measure on the borderlines.
- Comparative quantitative experiments
  - Measure and compare different systems
  - Remember: Easier to get results than to find the causes (causality)
- Experiments looking for **a** solution
  - Experiment until a solution has been found
  - Remember: Ought to be systematic to avoid trial and error

cphbusiness

Verification/falsification experiments example 1

Hyp

One instance don't have access to private attributes in other instances

Exp

Define a class that has a relation 'to it self' and a private attribute. Create 2 instances of the class and let one of these instances try to change the private attribute in the other instance

Exe

This was the result ….

Eva

Hereby we can conclude …

cphbusiness

Verification/falsification experiments example 2

Hyp

Import statements only have consequences for compile time and not for run time

Exp

Create a program and compile and run the program with import of all libraries and without import. Take the time with a stopwatch

Exe

This was the results ….

Eva

???

Generelly

cphbusiness

Hyp

Hypotheses shall be formulated so that it is possible to verify/falsify (expected results)

Exp

Experiments shall be described as setting up the experiment

Exe

Execution shall be described in form of concrete results (data)

Eva

Data from execution is related to the hypotheses (results (data) versus expected results (hypotheses))

Experiment iteration

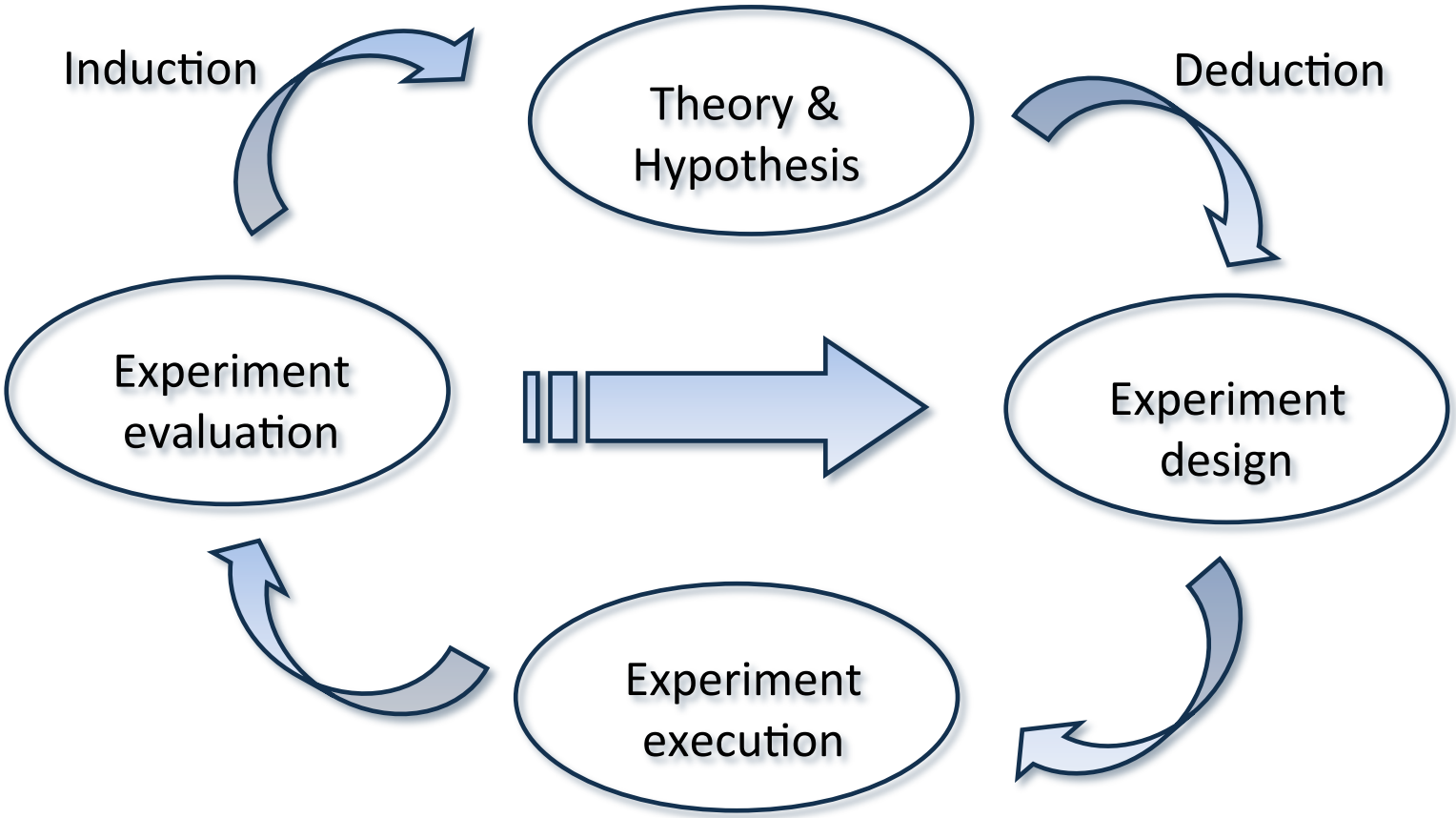
cphbusiness

- Often we need more that one experiment

- Iterations of experiments makes us learn (part of a learning process)
  - Experiments ends by defining the next relevant experiment
  - We switch between **inductive theories** (creation of the model) and **deductive hypotheses** (expectations based on the model)

Experiment sequences

cphbusiness



Complications

- No systems are closed systems

- **Define the system** under investigation
- Make a description of all the **factors** that might effect the system
- Define a **model of causality**
  - Which factors have which effects?
- Determine one **factor's effect** by changing this factor while holding all the other factors stable

Complications Example

cphbusiness

- Define the system:
  - Finding nodes in a list representation of a graph
- Factors:
  - Algorithm, data structure, data, JVM, the physical machine (MHz, ram), the network, the operation system
- Model of causality:
  - ???
- Factor effect:
  - Two different algorithms are used to look for the same nodes in the same graph on the same machine ….
- Evaluation:
  - Algorithm A is fastest, but what if we changes the graph?

Complications 3

- It shall be possible to redo the experiment – so…

- Never use random number
  - Use data from files
- Make a detailed description on the experiment's set up
  - Definition of the system, the systems relations to it surroundings
- Be aware that measuring – the experiment – can effect the results
  - i.e. System.out.println() or writing to a file costs time

Experiment ☺

cphbusiness

- Create a concrete example, including:
  - Why is it a relevant experiment?
  - Which type of experiment is it?
    - There are different types of experiments and it is important to know which types you work with and when
  - Description of the experiment by:
    - Model/theory
    - Hypothesis/ assumption
    - Experimental setup
    - Result
    - Evaluation

Kasper's bud på nogle cases

cphbusiness

- Jeg læste et sted (*det gjorde jeg så faktisk ikke*) at Java's HashMap viste sig ikke at være effektiv når dens nøgler var på formen "Hans-X", hvor X er et heltal. Specielt skulle der være problemer når man får mere end 1 million nøgler. Jeg er ikke sikker på om det var ved indsættelse eller opslag der var problemet. Eftersom vi havde regnet med at bruge det som en central datastruktur, så skal vi lige have be/afkræftet dette.

- Vi får besøg af nogle potentielle kunder fra spanien næste fredag. Kan vi lave en test database med 5.000.000 spaniere i, så det ser lidt rigtigt ud. Bynavne, for & efternavne, gade navne og adresser. Hvis vi kan få deres alder til at være fordelt så det passer med alders fordelingen i spanien.

- Vi har fået nogle data på csv format som vi skal have ind i databasen. Det er lidt uklart hvor lang tid vi skal afsætte til det? Det burde være simpelt, men jeg er altid lidt mistænksom.
  (Datafil)