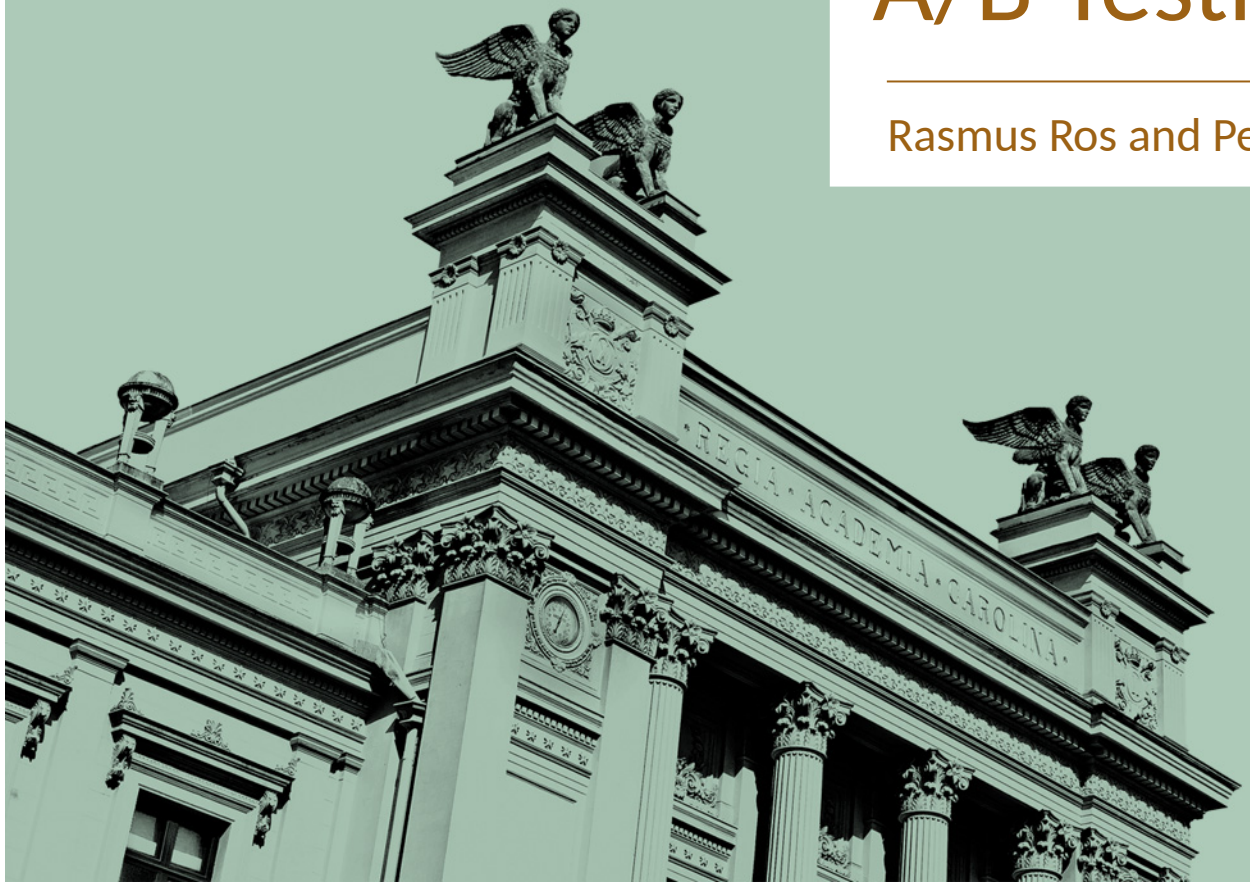
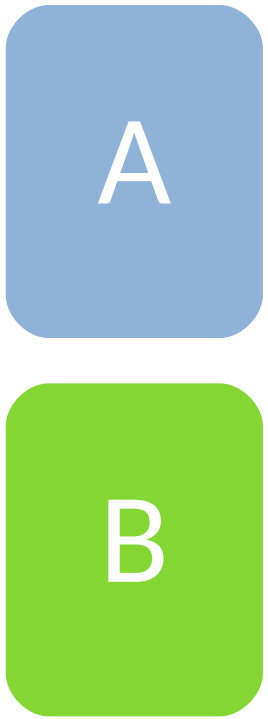
Continuous Experimentation and A/B Testing: A Mapping Study

Rasmus Ros and Per Runeson



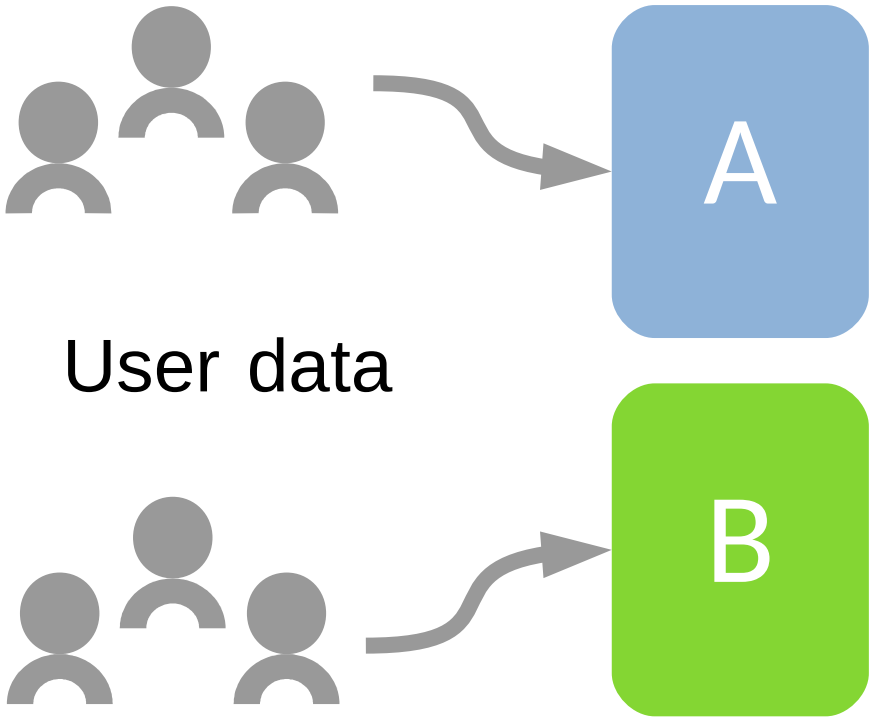
A/B Testing



- Software changes
  - User interface tweaks
  - Algorithm parameters
  - ”Old” vs ”new” feature

LUND UNIVERSITY

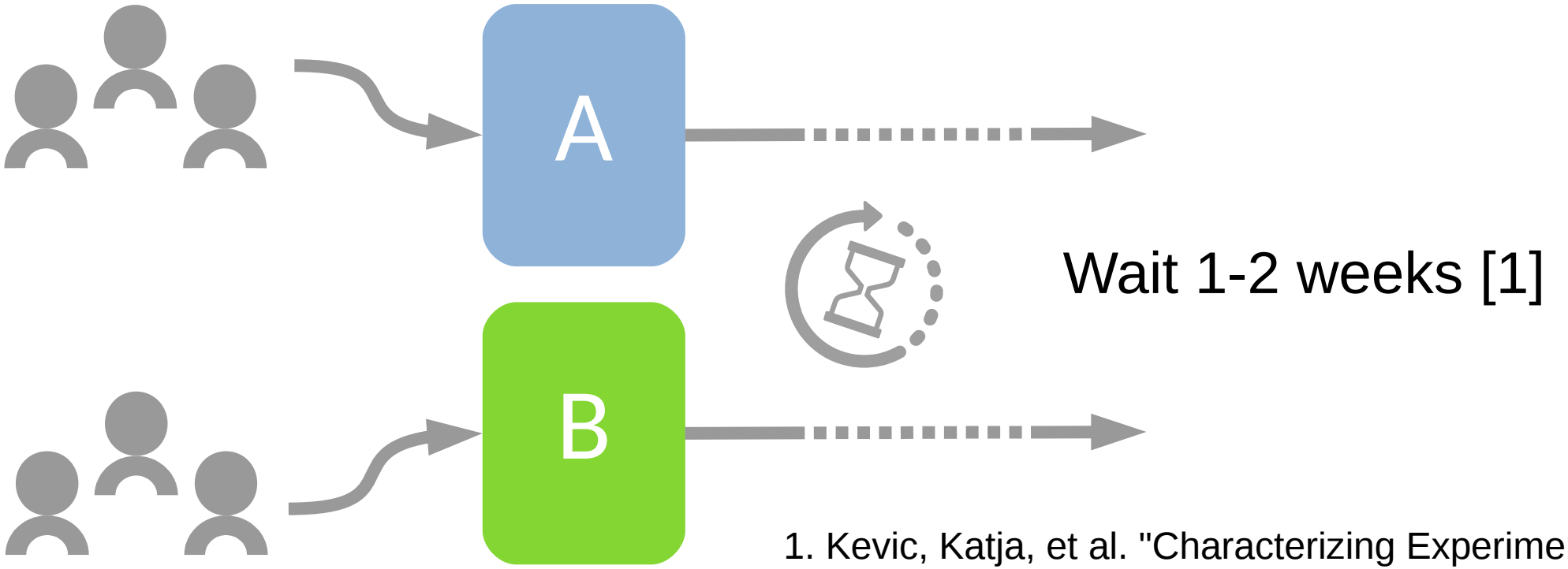
A/B Testing



http://bit.ly/rcose18ab

LUND UNIVERSITY

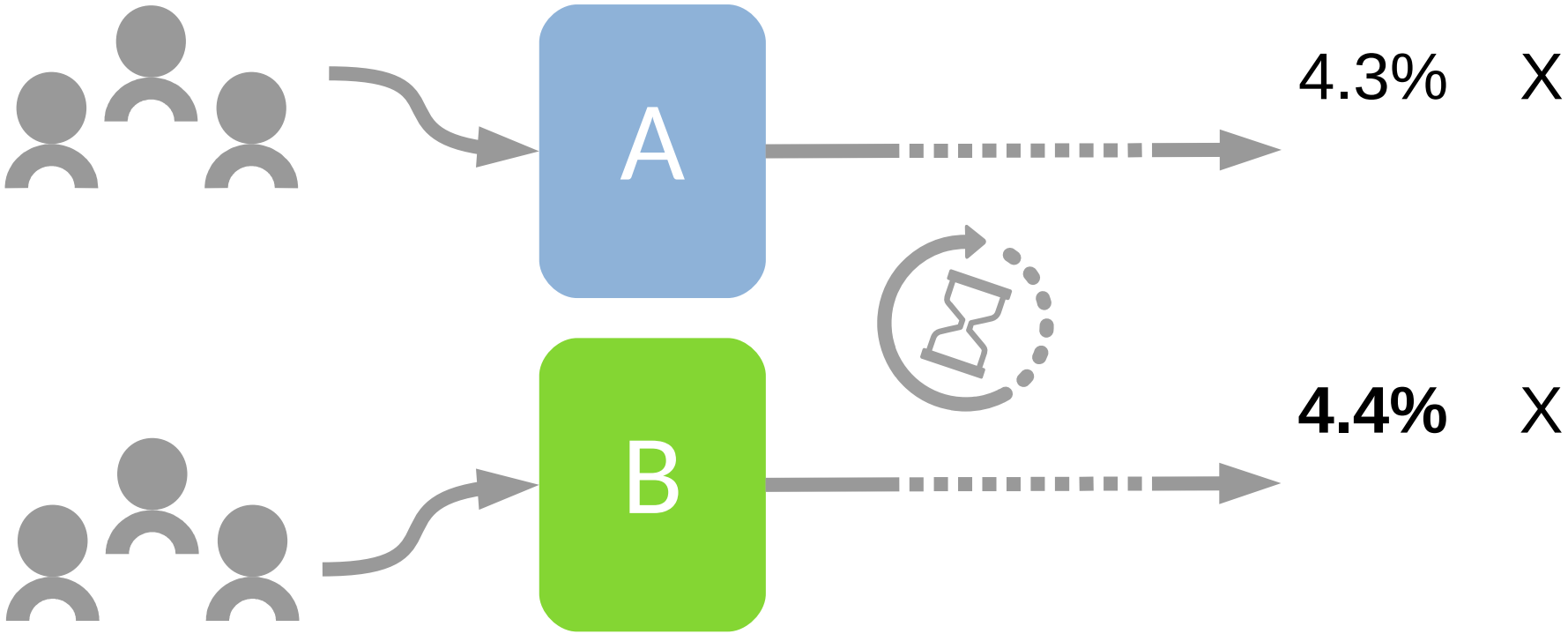
A/B Testing



1. Kevic, Katja, et al. "Characterizing Experimentation in Continuous Deployment: A Case Study on Bing." ICSE: SEIP. 2017. doi: 10.1109/ICSE-SEIP.2017.19

LUND UNIVERSITY

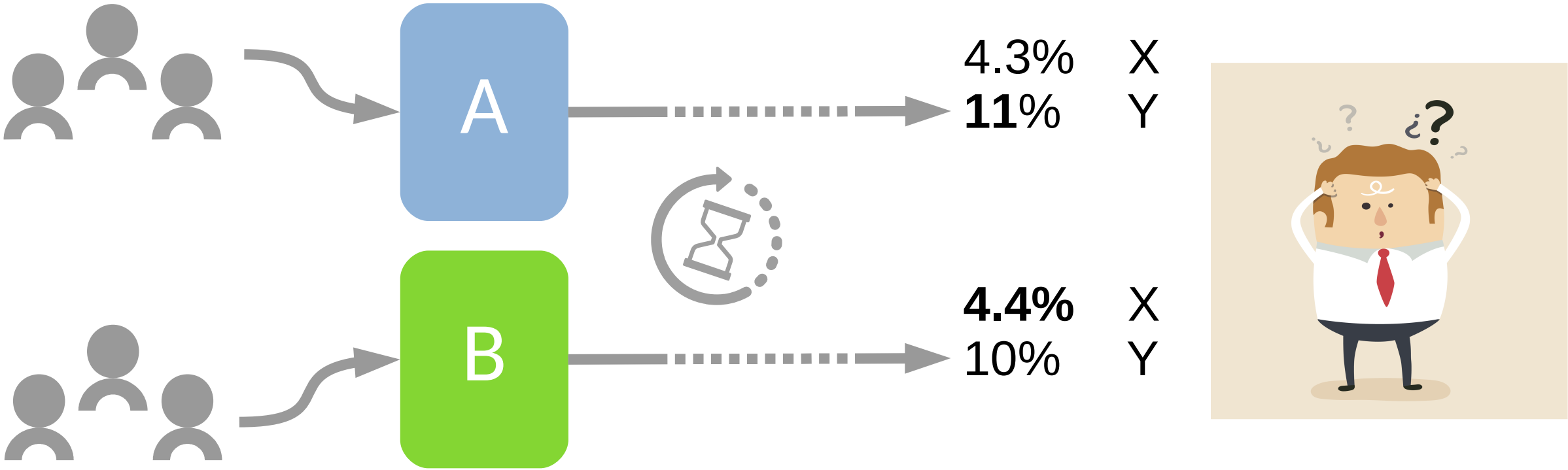
A/B Testing



 http://bit.ly/rcose18ab

LUND UNIVERSITY

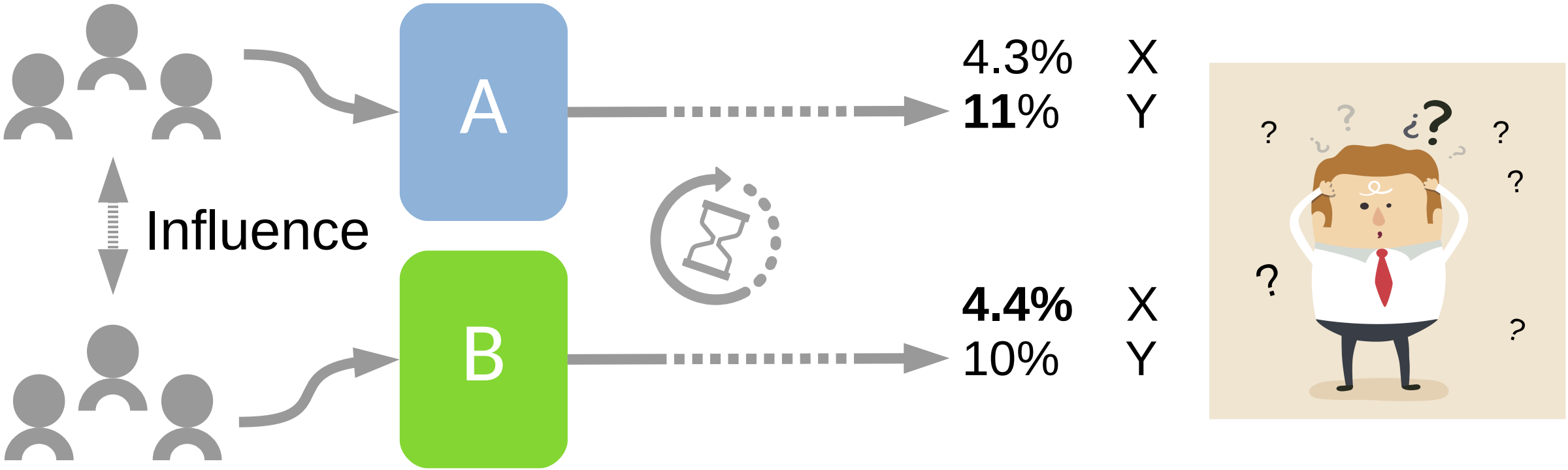
A/B Testing



 http://bit.ly/rcose18ab

LUND UNIVERSITY

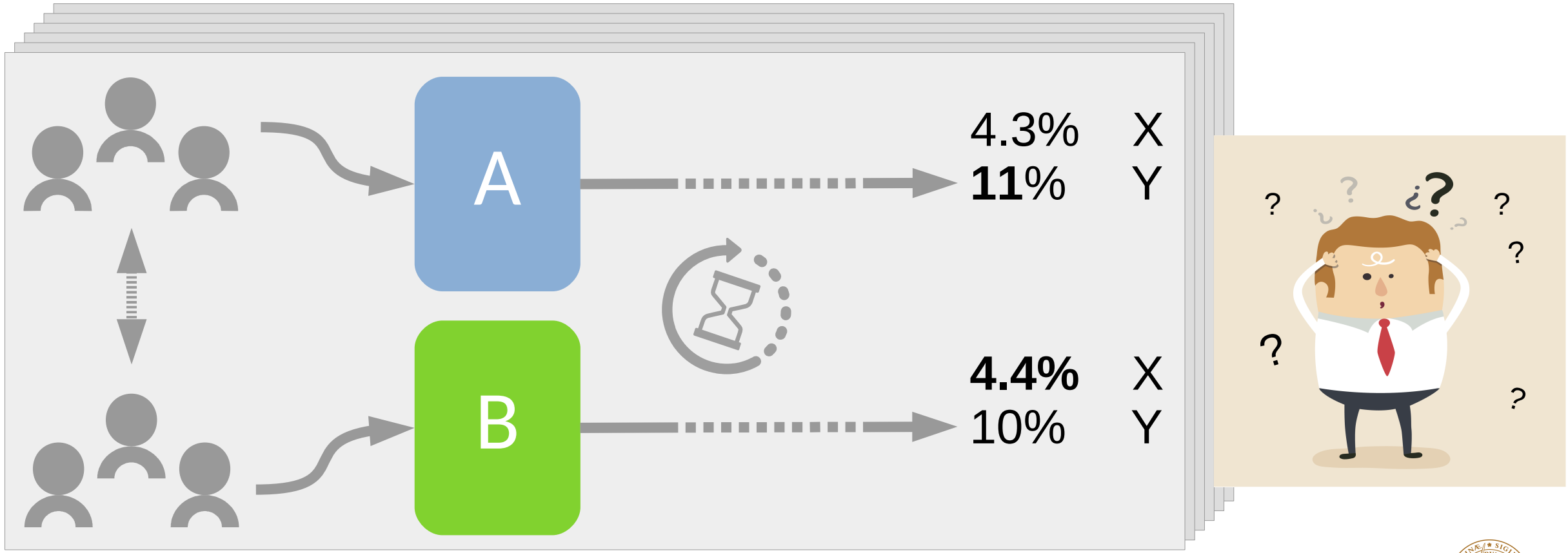
A/B Testing



 http://bit.ly/rcose18ab

LUND UNIVERSITY

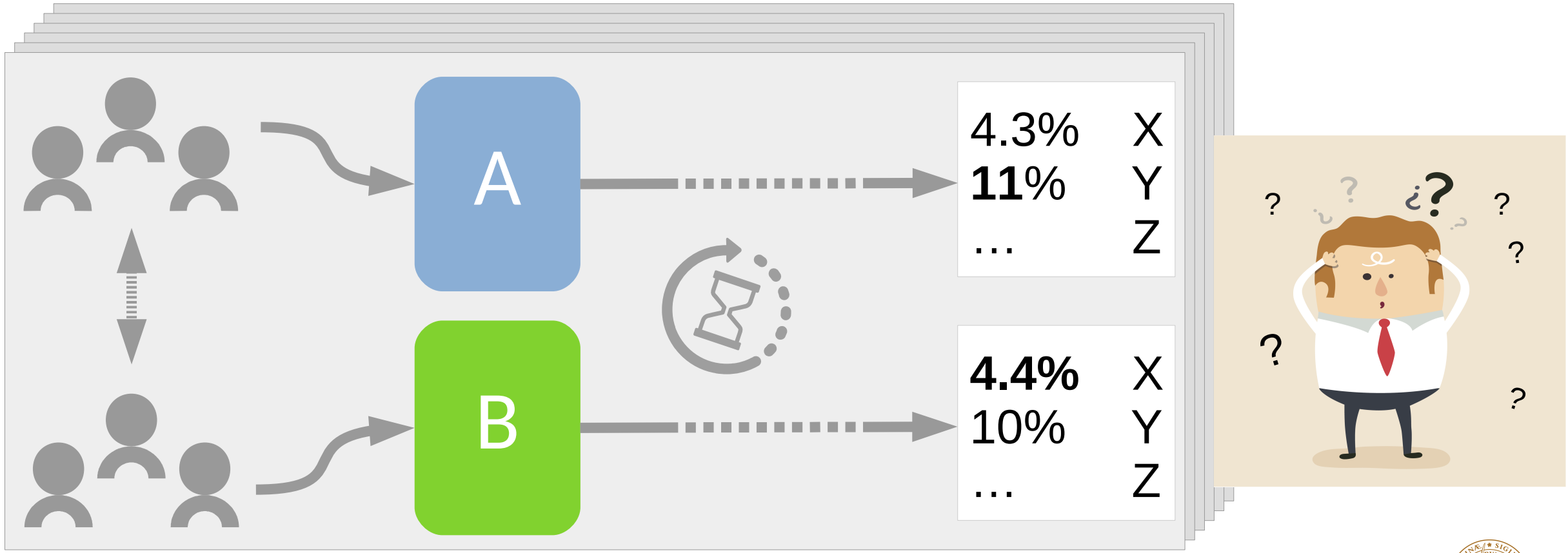
A/B Testing



http://bit.ly/rcose18ab

LUND UNIVERSITY

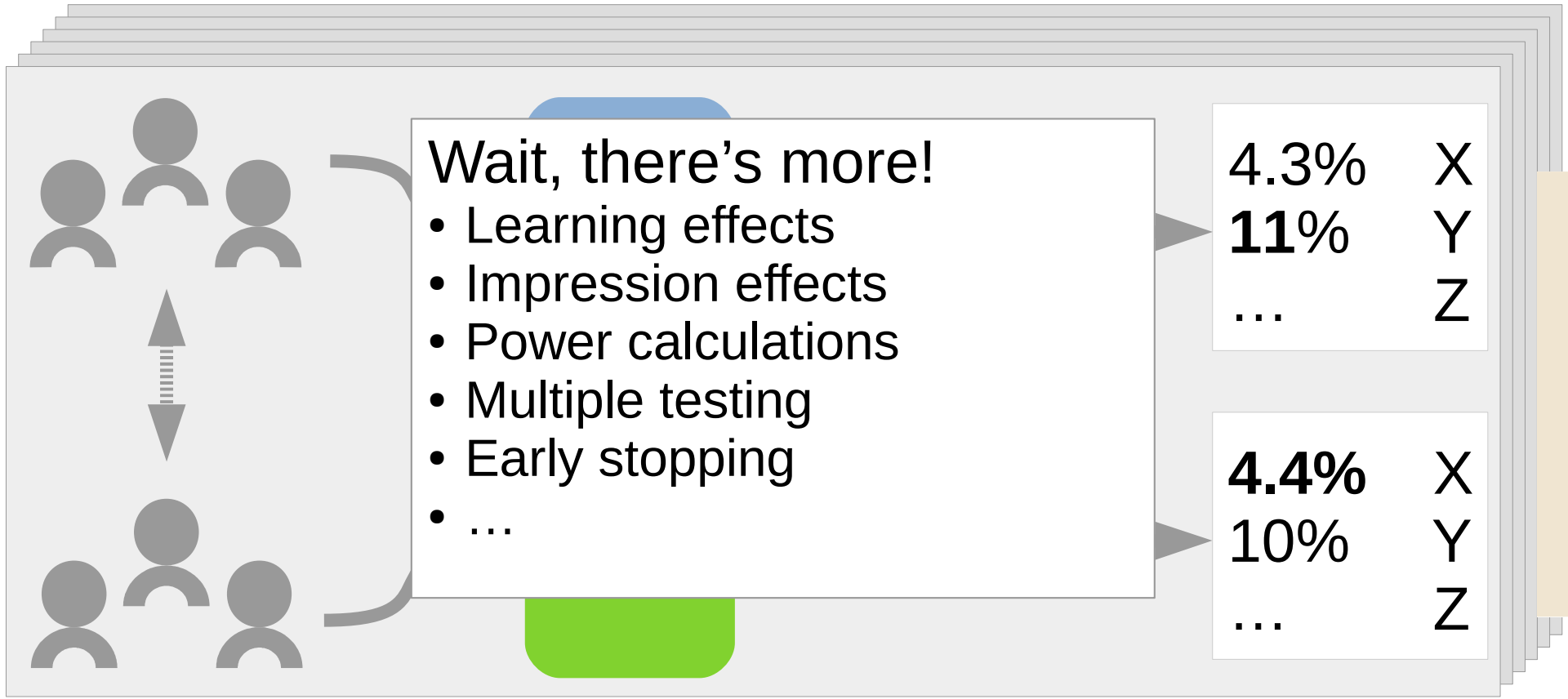
A/B Testing



http://bit.ly/rcose18ab

LUND UNIVERSITY

A/B Testing

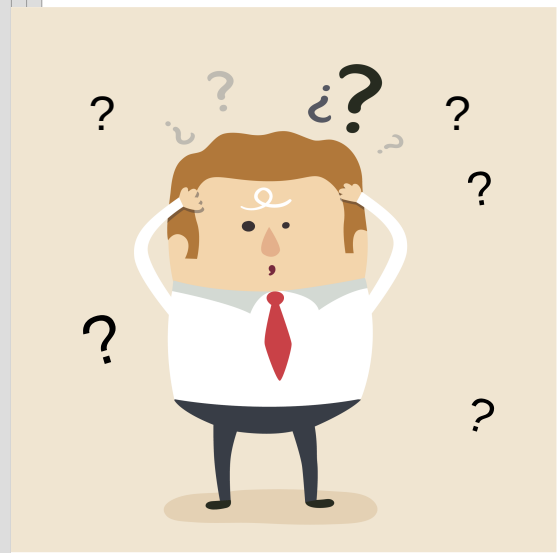


Wait, there's more!

- Learning effects
- Impression effects
- Power calculations
- Multiple testing
- Early stopping
- …

4.3%	X
11%	Y
…	Z

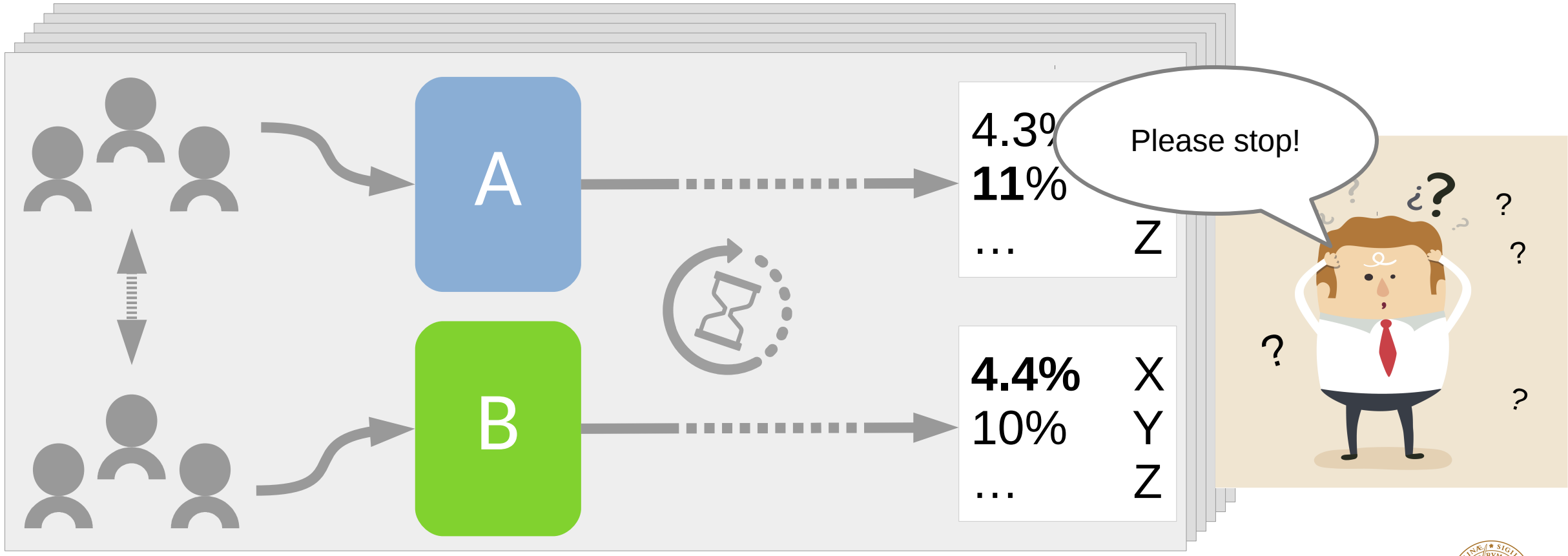
4.4%	X
10%	Y
…	Z



http://bit.ly/rcose18ab

LUND UNIVERSITY

A/B Testing



http://bit.ly/rcose18ab

LUND UNIVERSITY

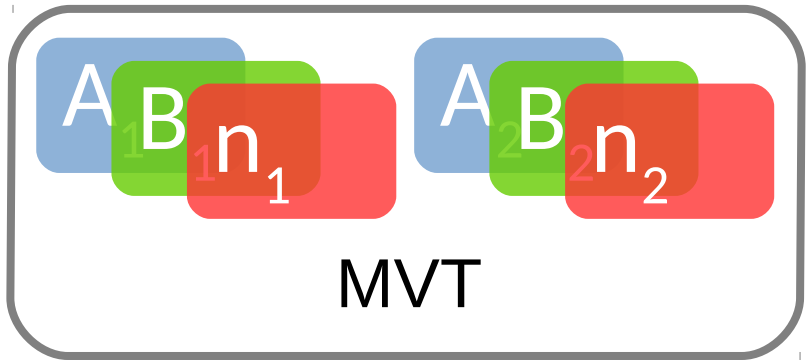
A/B Test

- Controlled experiment
  - 2 groups
  - Metric
  - Hypothesis test
- Also A/B/n test and MVT

LUND UNIVERSITY

A/B Test

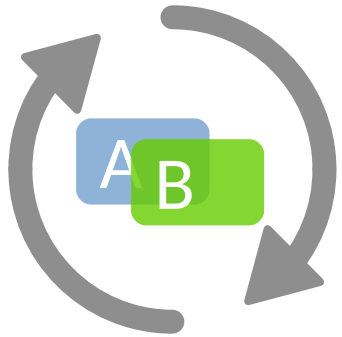
- Controlled experiment
  - 2 groups
  - Metric
  - Hypothesis test
- Also A/B/n test and MVT



LUND UNIVERSITY

Continuous Experimentation

- Continuous iterative process [2]
  1) Vision
  2) Business goals
  3) *Experiment*
  4) Learnings
- Roles, architecture, infrastructure, …
- Synergies with continuous *

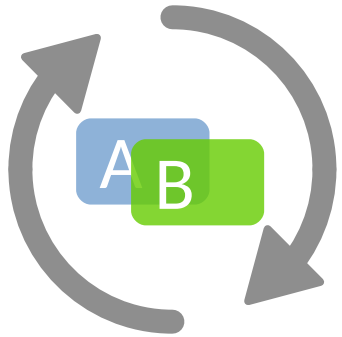


2. Fagerholm, Fabian, et al. "The RIGHT Model for Continuous Experimentation." Journal of Systems and Software 123 (2017): 292-305.

LUND UNIVERSITY

Continuous Experimentation

- Continuous iterative process [2]
  1) Vision
  2) Business goals
  3) *Experiment*
  4) Learnings
- Roles, architecture, infrastructure, …
- Synergies with continuous *

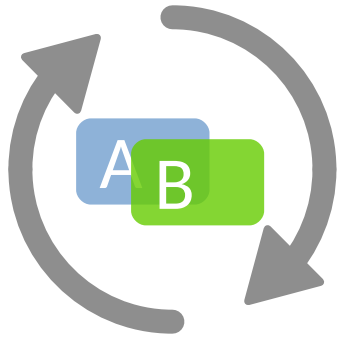


2. Fagerholm, Fabian, et al. "The RIGHT Model for Continuous Experimentation." Journal of Systems and Software 123 (2017): 292-305.

LUND UNIVERSITY

Continuous Experimentation

- Continuous iterative process [2]
  1) Vision
  2) Business goals
  3) *Experiment*
  4) Learnings
- Roles, architecture, infrastructure, …
- Synergies with continuous *



2. Fagerholm, Fabian, et al. "The RIGHT Model for Continuous Experimentation." Journal of Systems and Software 123 (2017): 292-305.

LUND UNIVERSITY

Overview

- **Aim** — State of research; Applicability of CE
- **Method** — Phrase search and references; Thematic analysis
- **Result** — 62 papers

LUND UNIVERSITY

Overview

- **Aim** — State of research; Applicability of CE
- **Method** — Phrase search and references; Thematic analysis
- **Result** — 62 papers

LUND UNIVERSITY

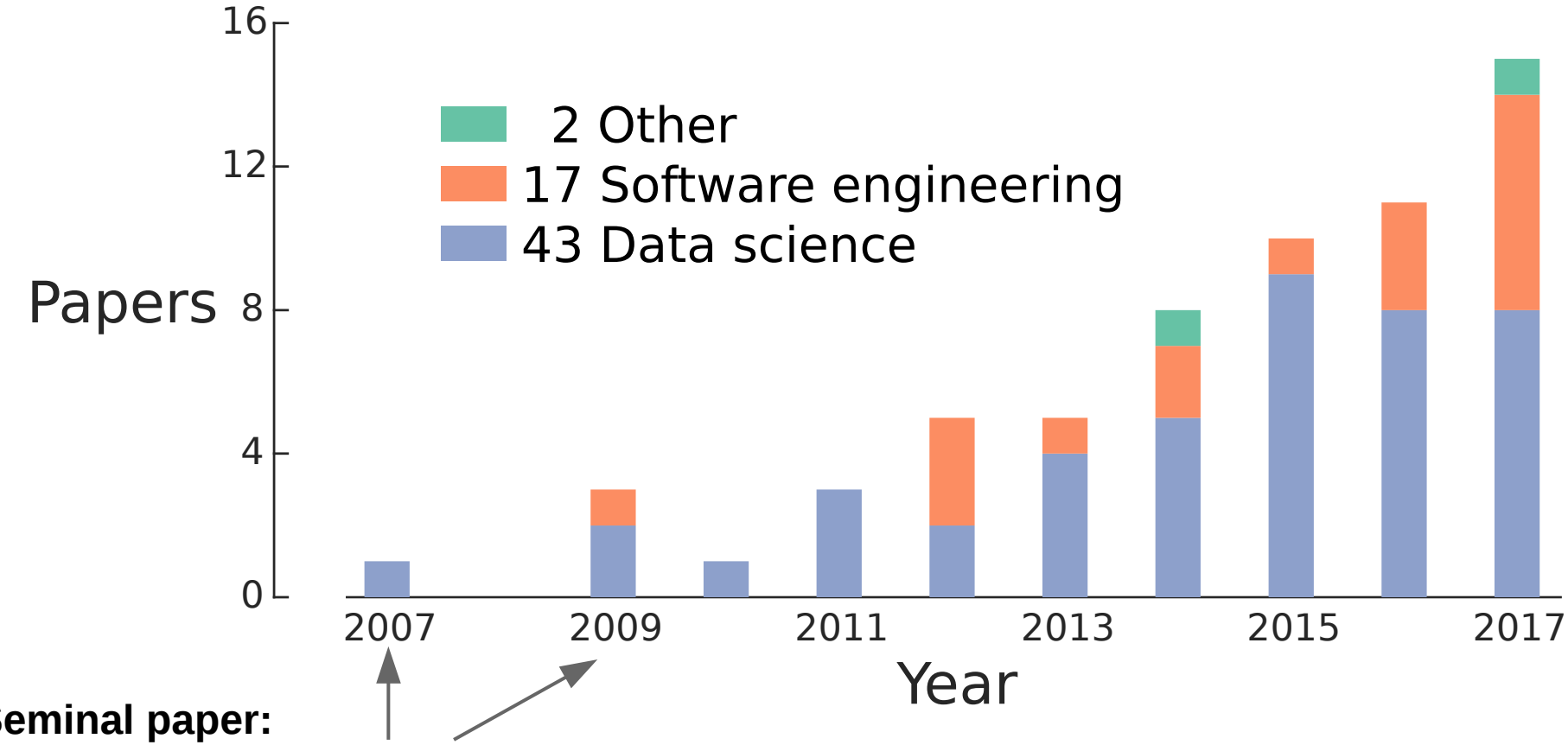
Overview

- **Aim** State of research
  Applicability of CE
- **Search** Phrase search and references
  Thematic analysis
- **Result** 62 papers

http://lup.lub.lu.se/search/ws/files/40009496/extracted_data.csv

LUND UNIVERSITY

Research Trend



Seminal paper:

3. Kohavi, Ron, et al. "Controlled Experiments on the Web: Survey and Practical Guide." Data Mining and Knowledge Discovery (SIGKDD) (2009).

LUND UNIVERSITY

Research Questions

- **RQ1** What are the main topics researched within CE and how are they studied?
- **RQ2** Which kind of organizations use CE and which sectors do they operate in?
- **RQ3** With what type of experiments have CE been applied?

LUND UNIVERSITY

Research Questions

- **RQ1** What are the main topics researched within CE and how are they studied?
- **RQ2** Which kind of organizations use CE and which sectors do they operate in?
- **RQ3** With what type of experiments have CE been applied?

LUND UNIVERSITY

Research Questions

- **RQ1** What are the main topics researched within CE and how are they studied?
- **RQ2** Which kind of organizations use CE and which sectors do they operate in?
- **RQ3** With what type of experiments have CE been applied?

LUND UNIVERSITY

Research Topics (RQ1)

Topic	Total
Experiment process	7
Infrastructure	10
Challenges	19
Benefits	3
Variability management	5
Metrics	6
Statistical methods	16
Design of experiments	8
Domain considerations	6
Ethics	1

LUND UNIVERSITY

Research Topics (RQ1)

Topic	Total
Experiment process	7
Infrastructure	10
Challenges	19
Benefits	3
Variability management	5
Metrics	6
Statistical methods	16
Design of experiments	8
Domain considerations	6
Ethics	1

- Evaluation
  - Evaluation research
  - Experience report
- Solution
  - Validation
  - Proposed solution

LUND UNIVERSITY

Research Topics (RQ1)

		Research approach	
Topic	**Total**	Evaluation	Solution
Experiment process	7	7	0
Infrastructure	10	8	2
Challenges	19	17	2
Benefits	3	3	0
Variability management	5	0	5
Metrics	6	3	3
Statistical methods	16	1	15
Design of experiments	8	2	6
Domain considerations	6	3	3
Ethics	1	1	0

LUND UNIVERSITY

Research Topics (RQ1)

		Research approach	
Topic	**Total**	Evaluation	Solution
Experiment process	7	7	0
Infrastructure	10	8	2
Challenges	**19**	**17**	**2**
Benefits	**3**	**3**	**0**
Variability management	5	0	5
Metrics	6	3	3
Statistical methods	16	1	15
Design of experiments	8	2	6
Domain considerations	6	3	3
Ethics	1	1	0

- Anna Karenina principle

LUND UNIVERSITY

Research Topics (RQ1)

		Research approach	
Topic	**Total**	Evaluation	Solution
Experiment process	7	7	0
Infrastructure	10	8	2
Challenges	19	17	2
Benefits	3	3	0
Variability management	**5**	**0**	**5**
Metrics	6	3	3
Statistical methods	**16**	**1**	**15**
Design of experiments	**8**	**2**	**6**
Domain considerations	6	3	3
Ethics	**1**	**1**	**0**

- Technical topics
- Ethics guidelines

LUND UNIVERSITY

Research Topics (RQ1)

		Research approach	
Topic	**Total**	Evaluation	Solution
Experiment process	7	7	0
Infrastructure	10	8	2
Challenges	19	17	2
Benefits	3	3	0
Variability management	**5**	**0**	**5**
Metrics	**6**	**3**	**3**
Statistical methods	**16**	**1**	**15**
Design of experiments	**8**	**2**	**6**
Domain considerations	6	3	3
Ethics	1	1	0

- No data sets
- One open source tool

LUND UNIVERSITY

Organizations – Where is A/B Testing Used? (RQ2)

- Sectors
- Business model
- Company size

Subscribed or free	**Σ 29**
E-commerce	9
Search engine	5
Other	15

Perpetual	**Σ 9**
Finance	4
Gaming	2
Other	4

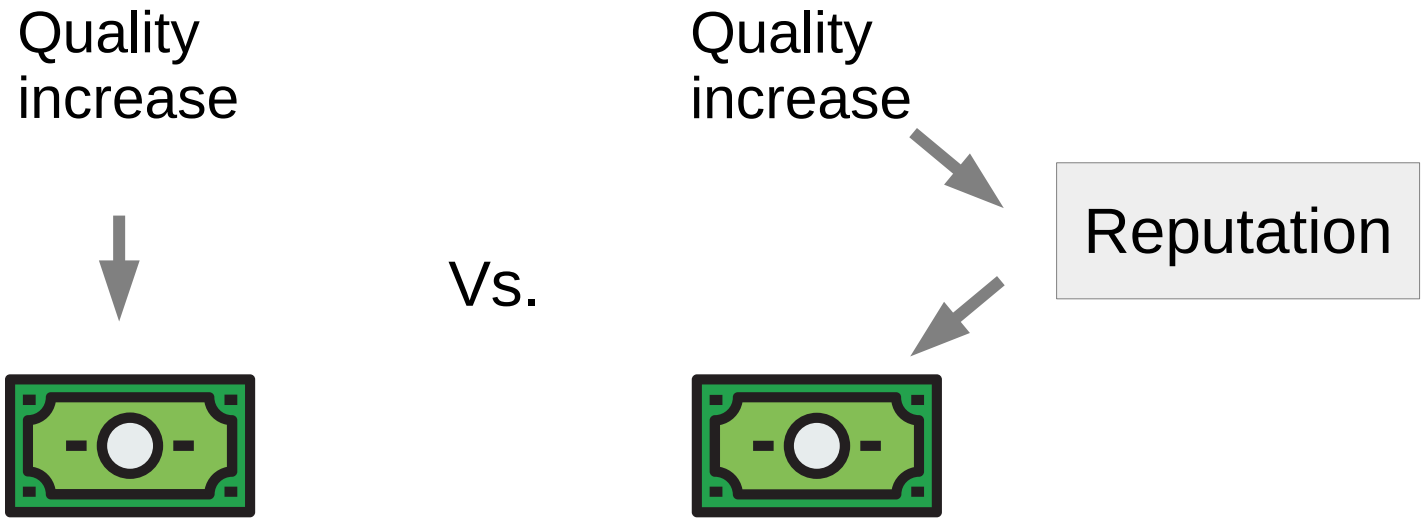
Embedded *	**Σ 4**

LUND UNIVERSITY

Organizations – Where is A/B Testing Used? (RQ2)

- Sectors
- Business model
- Company size

Business model	
Business to consumer (B2C)	32
Business to business (B2B)	10



LUND UNIVERSITY

Organizations – Where is A/B Testing Used? (RQ2)

- Sectors
- Business model
- Company size (employees)

Company size		
Large	≥ 250	29
Medium	< 250	5
Small	< 50	8

LUND UNIVERSITY

Experiments – What and How? (RQ3)

- Treatment
- Goal
- Experiment design

Treatment	
Visual change	64
Algorithmic change	23
New feature	4

LUND UNIVERSITY

Experiments – What and How? (RQ3)

- Treatment
- Goal
- Experiment design

Goal	
Engagement	58
Revenue	26
Knowledge	7

LUND UNIVERSITY

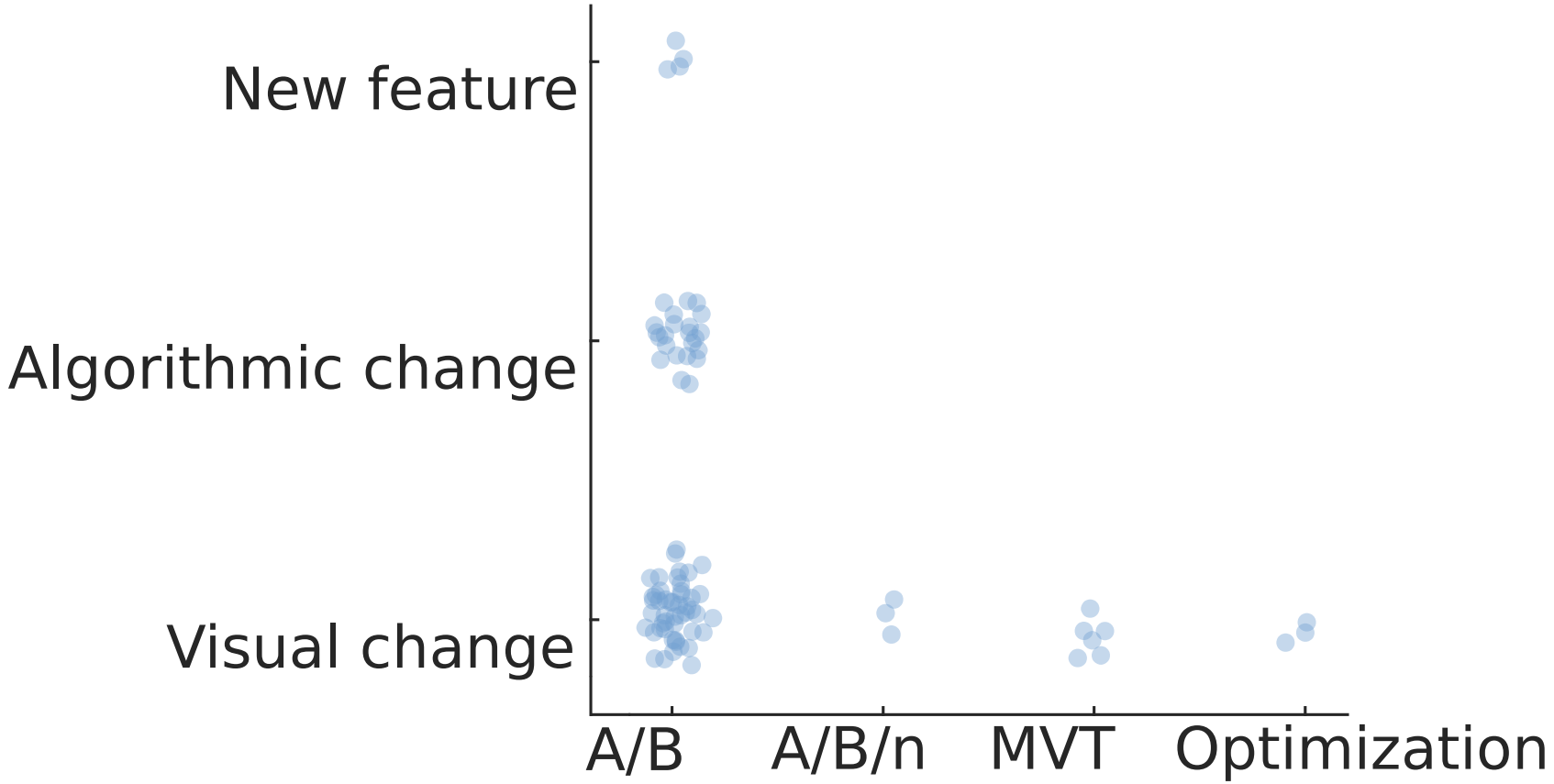
Experiments – What and How? (RQ3)

- Treatment
- Goal
- Experiment design

Experiment design	
A/B	77
A/B/n	3
MVT	6
Optimization	3

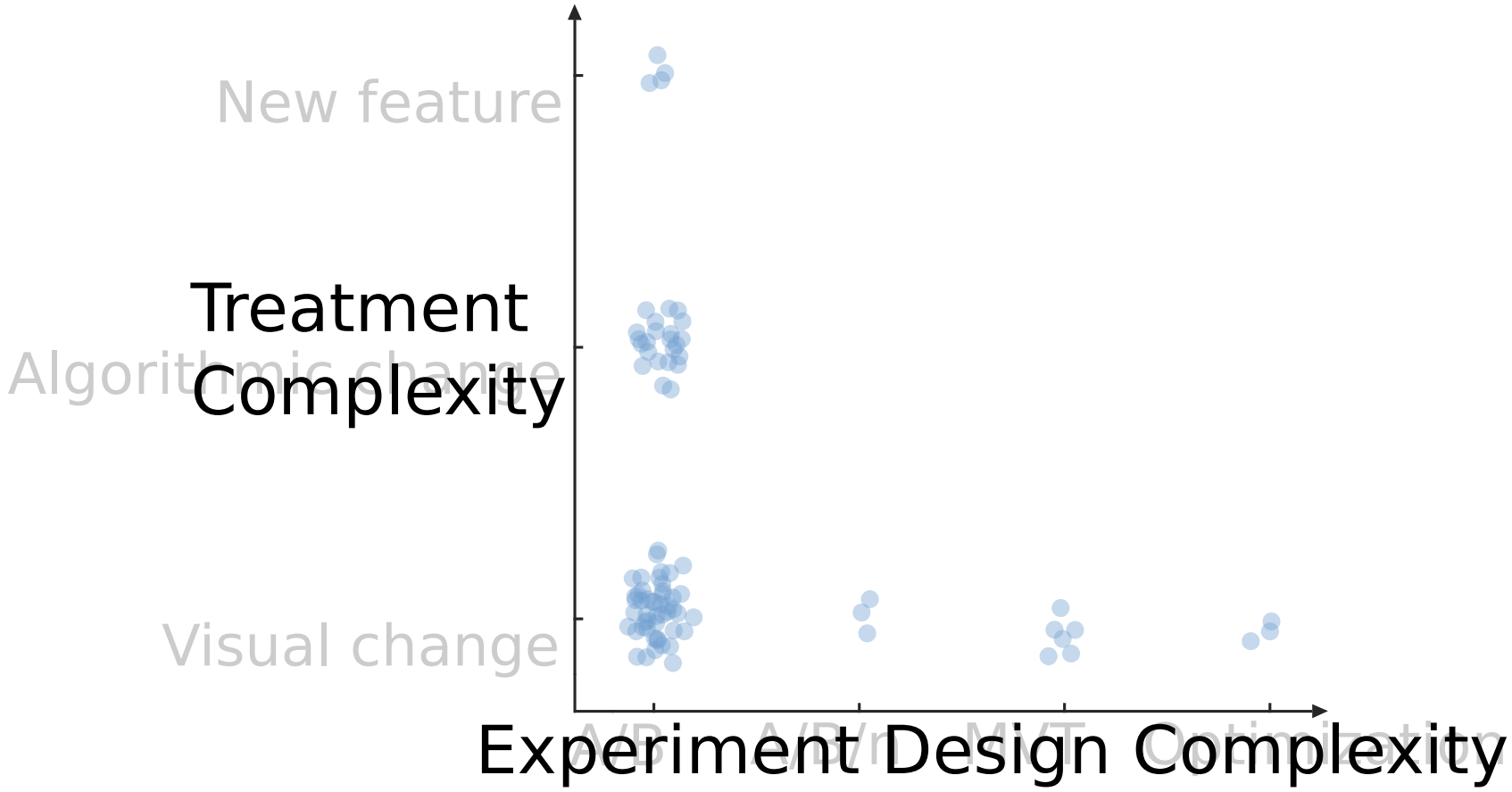
LUND UNIVERSITY

Slicing Experiments



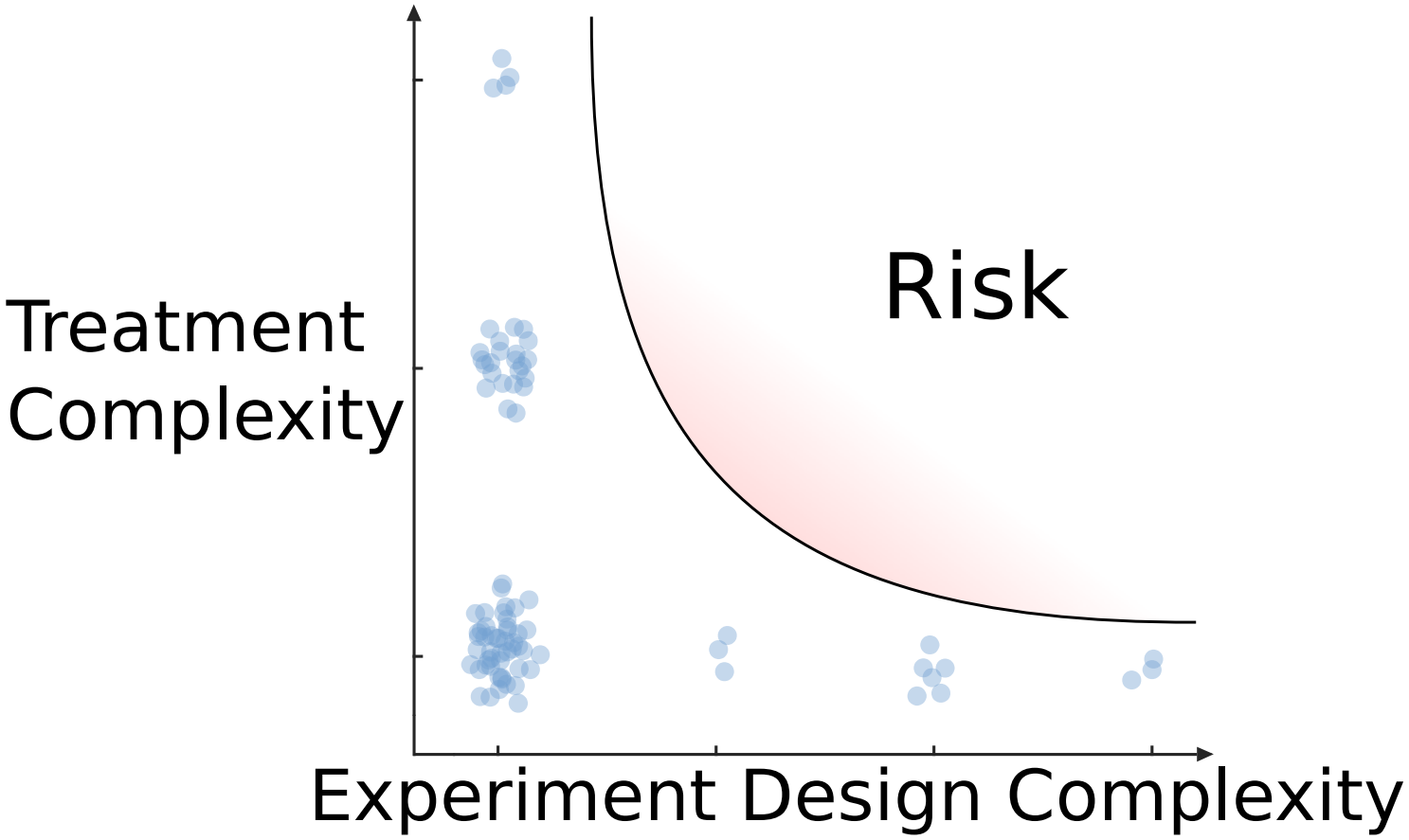
LUND UNIVERSITY

Slicing Experiments



LUND UNIVERSITY

Slicing Experiments



LUND UNIVERSITY

Take Aways

1) Research gaps
   - Tool evaluations
   - Ethics guidelines
   - Embedded

2) Diverse organisations do A/B testing

3) Simple experimental designs – (Best) practice or bias?

LUND UNIVERSITY

Take Aways

1) Research gaps
   - Tool evaluations
   - Ethics guidelines
   - Embedded

2) Diverse organisations do A/B testing

3) Simple experimental designs – (Best) practice or bias?

LUND UNIVERSITY

Take Aways

1) Research gaps
   - Tool evaluations
   - Ethics guidelines
   - Embedded

2) Diverse organisations do A/B testing

3) Simple experimental designs – (Best) practice or bias?

LUND UNIVERSITY

LUND
UNIVERSITY

WASP | WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

”Out-of-Scope” A/B Testing Highlights

- Qubit meta analysis
  - https://www.qubit.com/wp-content/uploads/2017/12/qubit-research-meta-analysis.pdf
- Errors in statistics
  - https://blog.sumall.com/journal/optimizely-got-me-fired.html
  - https://www.aarondefazio.com/tangentially/?p=83
- Continuous Experimentation origin
  - http://mcfunley.com/design-for-continuous-experimentation
- A/B testing superflous changes
  - https://iterativepath.wordpress.com/2012/10/29/testing-40-shades-of-blue-ab-testing/
  - http://stopdesign.com/archive/2009/03/20/goodbye-google.html
- Ethics of experiments
  - https://doi.org/10.1073/pnas.1320040111
  - https://freedom-to-tinker.com/2014/07/08/on-the-ethics-of-ab-testing/
  - http://observer.com/2014/07/okcupid-brags-about-manipulating-online-daters-in-secret-tests/

LUND UNIVERSITY