Process Miner, are you sure?

On reliability and validity in process mining research

Prof. Dr. Jana-Rebecca Rehse

ICPM 2023 Doctoral Consortium Keynote

Major credits go to my colleagues
Jan Martijn van der Werf,
Sander Leemans, and Peter Fettke

Process Mining Papers

It provides better results than the other approach!

It works on real-life data!

It can deal with large event logs!

It achieved an F-score above 0.8!

Our new (really cool!) process mining approach

1) Introduction
2) Foundations
3) Approach
   a) Step 1
   b) Step 2
   c) Step 3
4) Evaluation
5) Related Work
6) Discussion
7) Conclusion

How do you know that your process mining approach still works when…

- it is applied by someone else?
- it is applied in another setting?
- it is applied on new data?

Will your claims about the approach still hold?

A shift in process mining research

UNIVERSITY OF MANNHEIM
Business School

Historic development

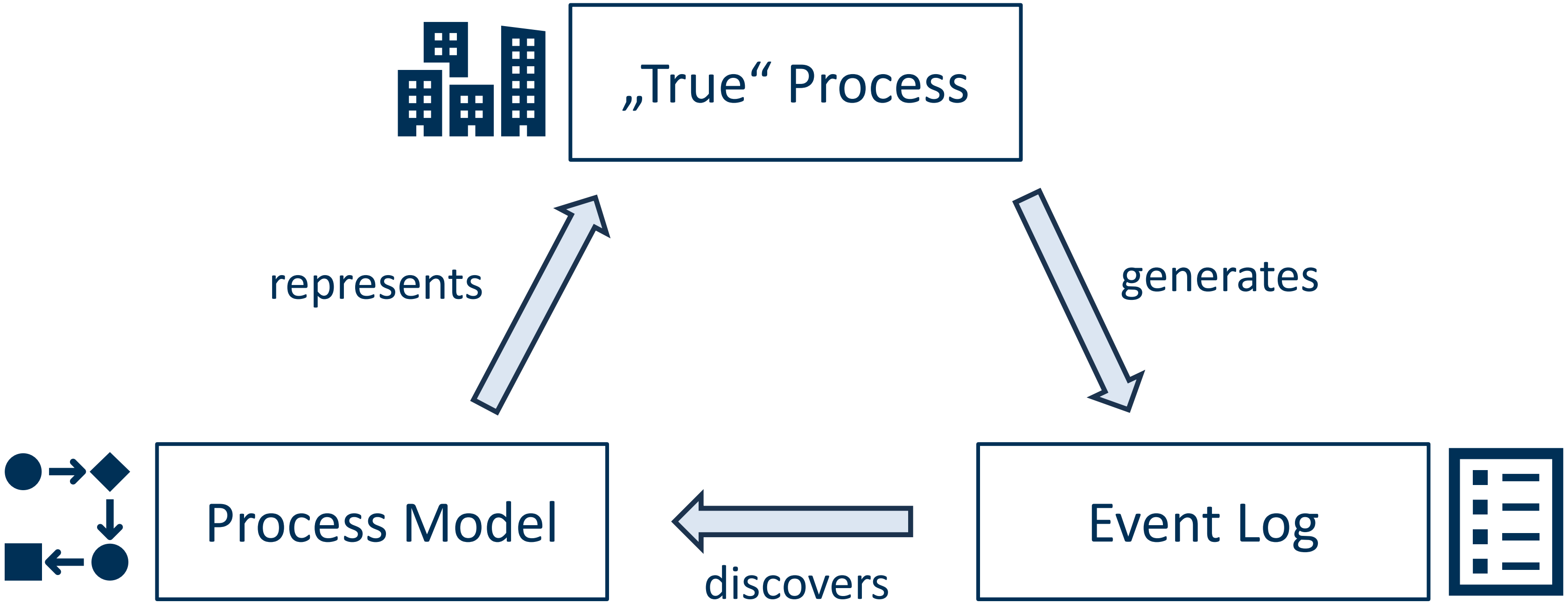
- Theoretical computer science, mathematical modeling, automata theory
- Typical results: properties of models and / or algorithms
- Typical methods: (formal) mathematical proofs

Current reality

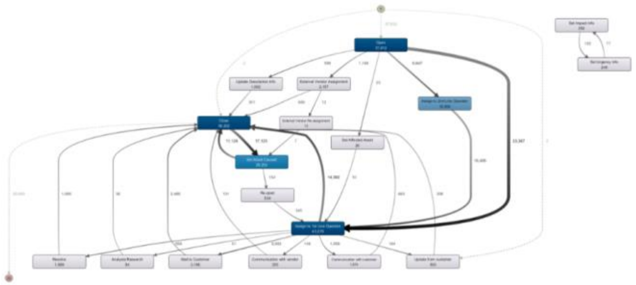
- Process mining is applied in organizations and practical settings
- Focus lies on data analysis
- Data is influenced by IT systems, people and other social constructs
- Properties of this data cannot be formally proven

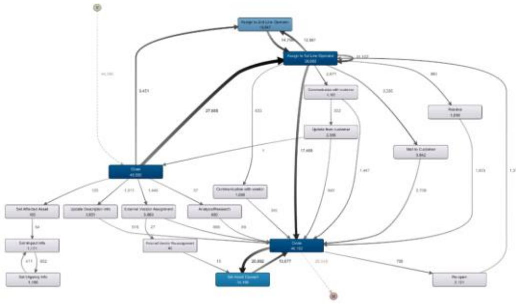
Conclusion: We need to use more **empirical** methods from social science instead of formal methods from mathematics or CS.

Process Discovery: Illustration



Process Discovery: Illustration





„True" Process

generates

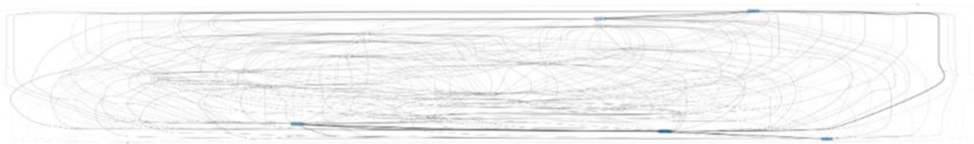
Event Log

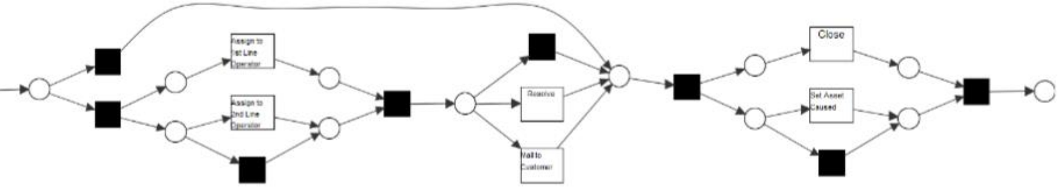
discovers

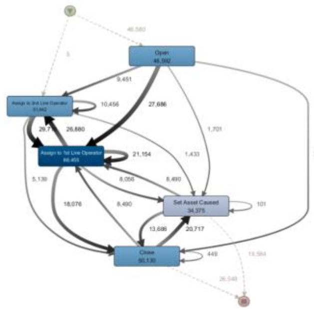
Process Model

represents

Which model represents the process best?

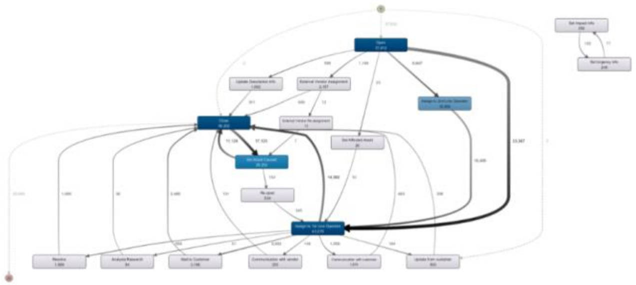


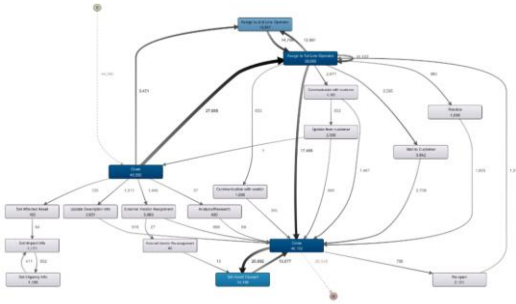




Process Discovery: Illustration

UNIVERSITY OF MANNHEIM
Business School





"able to replay event log"

replay fitness

"Occam's razor"

simplicity

process discovery

generalization

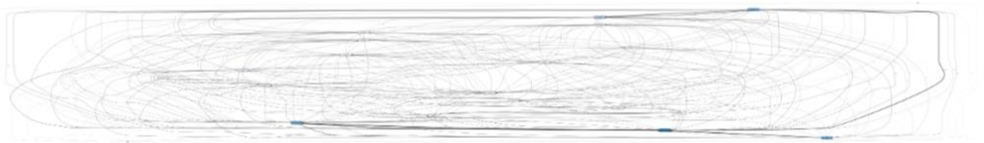
"not overfitting the log"

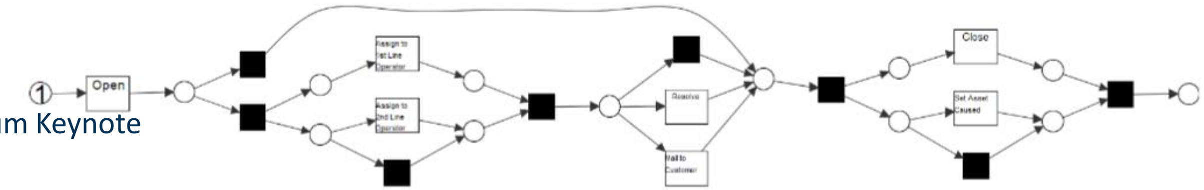
precision

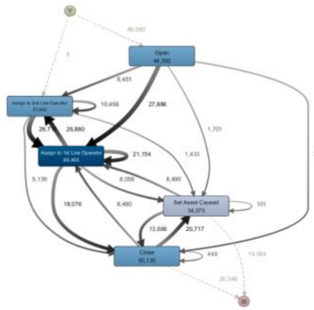
"not underfitting the log"

How should we weigh the different dimensions?

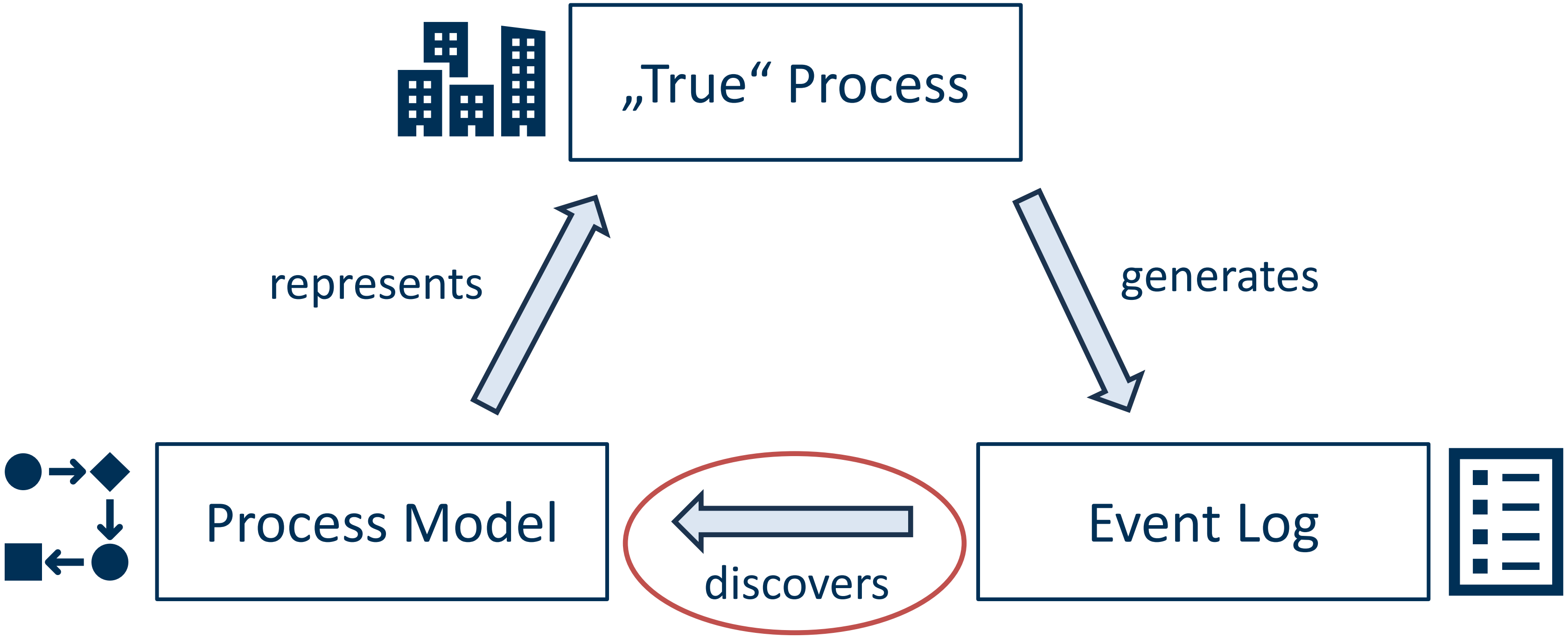
How can we measure them?



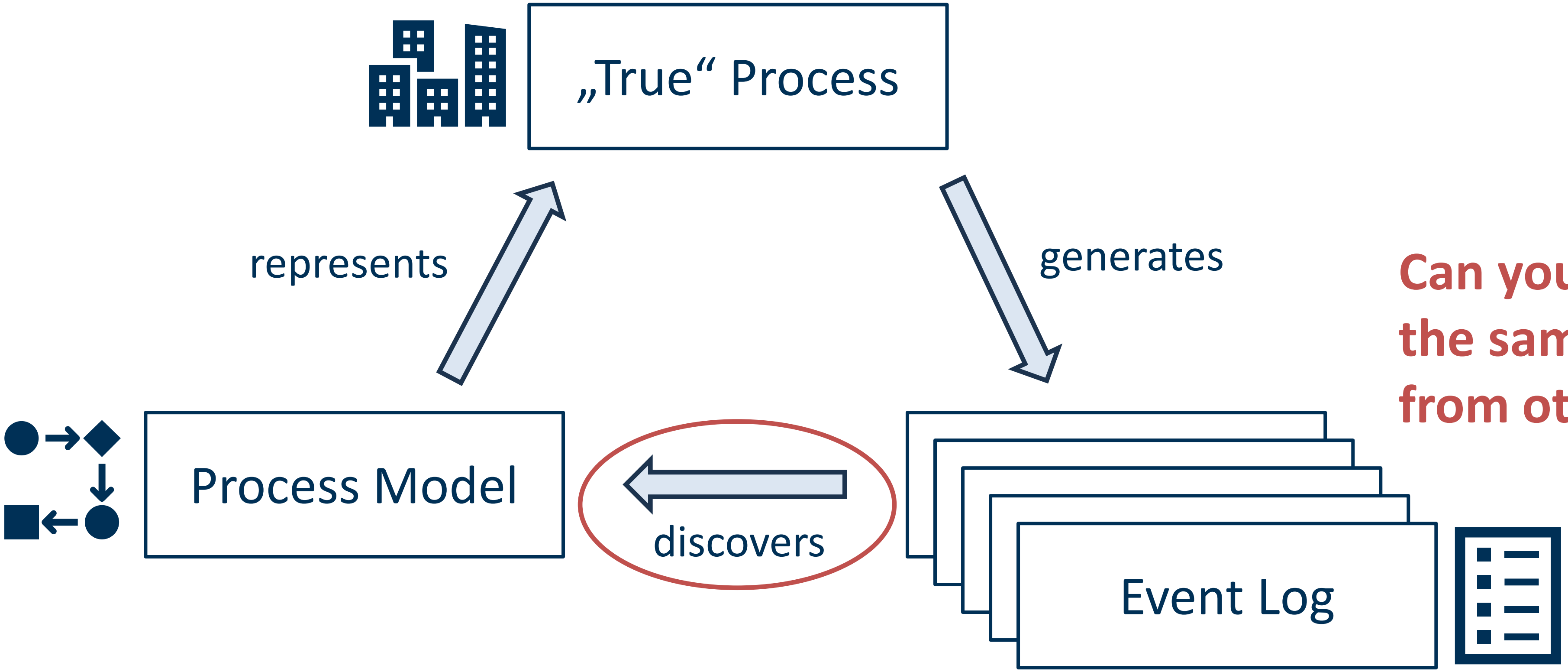




Process Discovery: Illustration

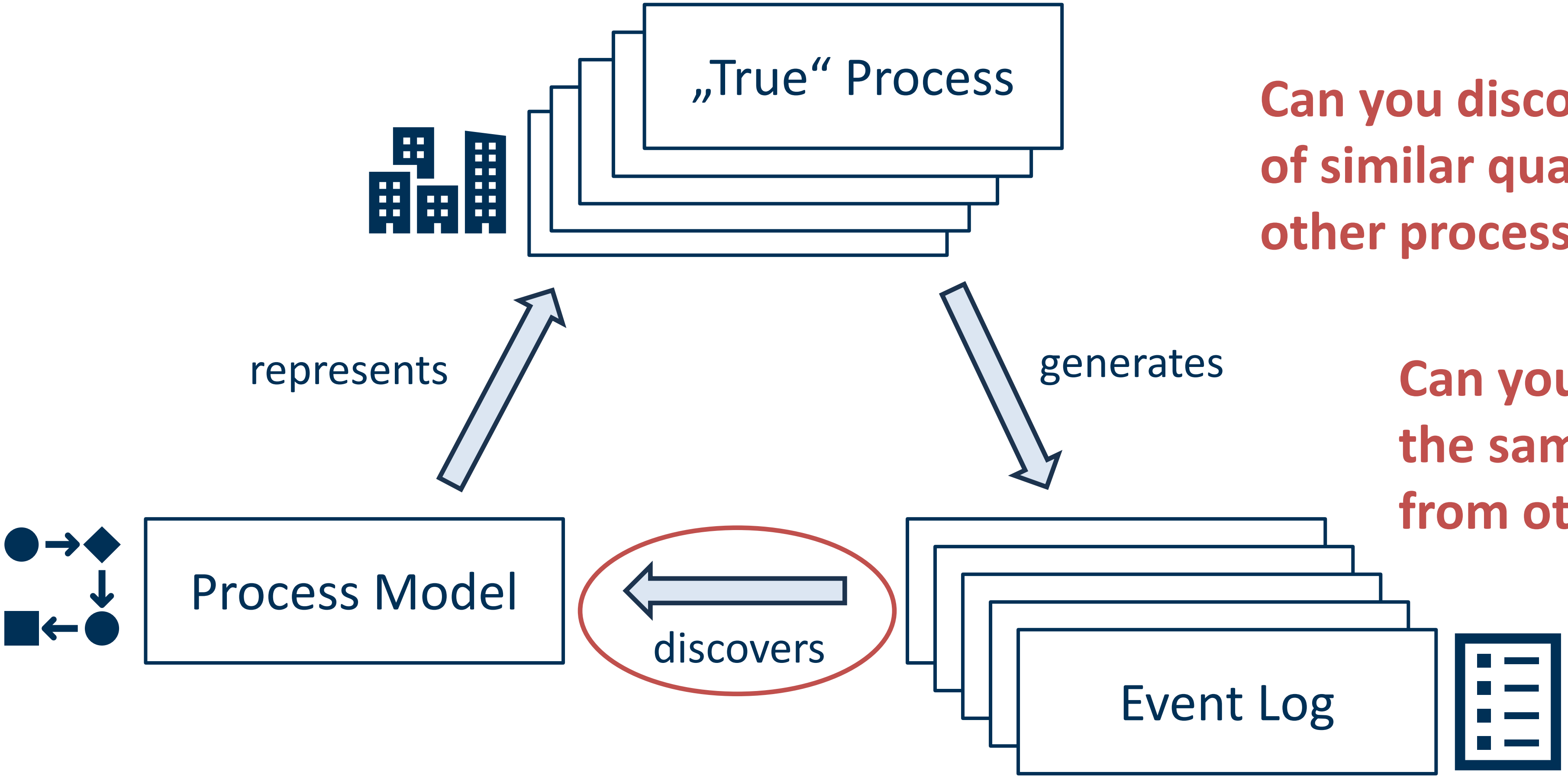


Process Discovery: Illustration



Can you discover the same process from other logs?

Process Discovery: Illustration



Can you discover models of similar quality for other processes?

Can you discover the same process from other logs?

Principles of scientific inquiry

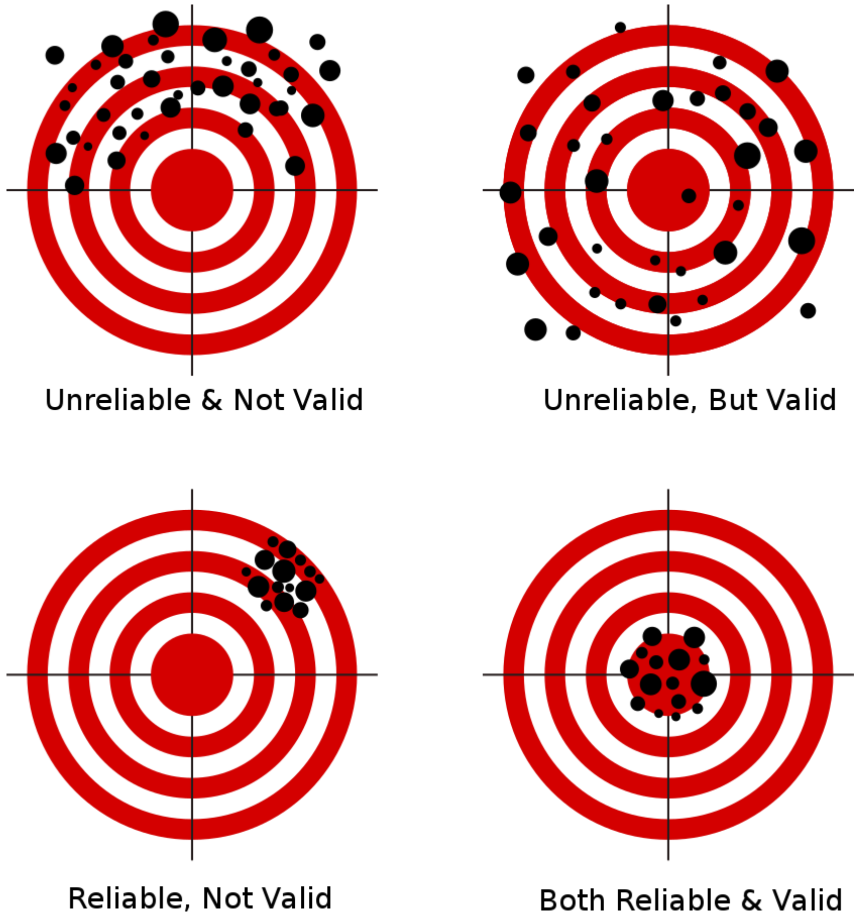
Degree to which a measure of a concept is stable

Reliability

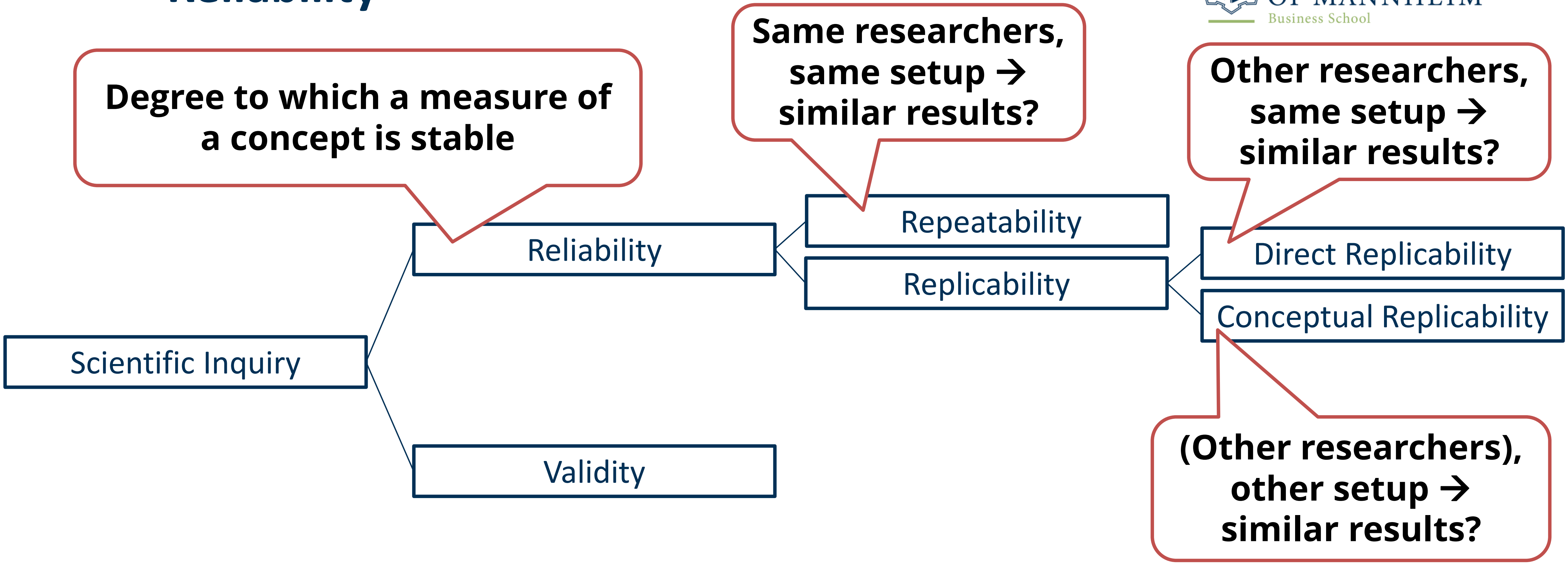
Scientific Inquiry

Validity

Degree to which results are free from errors



Reliability



Validity

Scientific Inquiry

Reliability

Validity

Degree to which results are free from errors

Conclusion Validity

Experiment allows for stated conclusions

Construct Validity

Measures assess the intended property

Soundness

Completeness

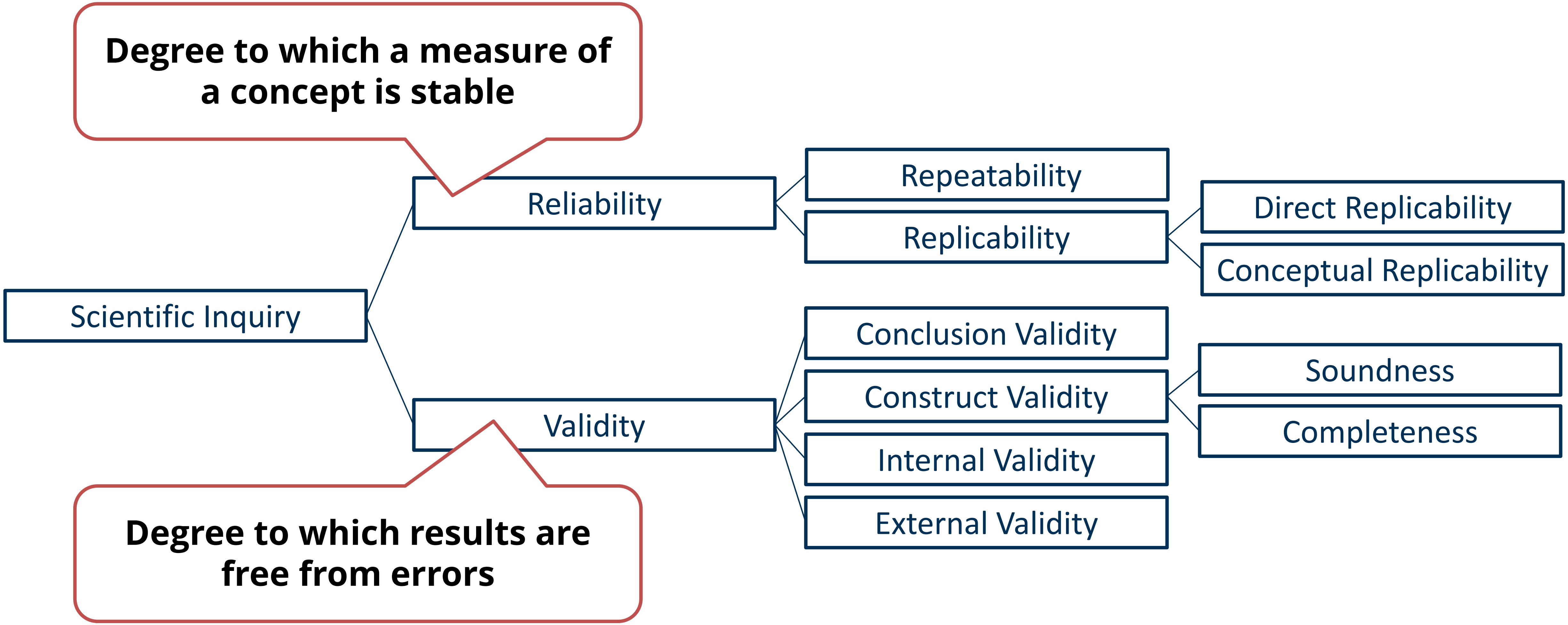
Internal Validity

Observed effects can be attributed to treatment

External Validity

Identified causalities also hold in other settings

What could possibly go wrong?



(Some) Process mining crimes

- Using the wrong evaluation data
  - E.g., overgeneralizing from “simplistic” logs (external validity)
- Misleading quality assessment
  - E.g., using selective measures (internal validity)
- Scientific inaccuracies
  - E.g., not evaluating all claims (construct validity)
- Improper comparison of results
  - E.g., improper treatment of competitors (conclusion validity)
- Missing information
  - E.g., only relative numbers (direct replicability)

Check for updates

Process Mining Crimes – A Threat to the Validity of Process Discovery Evaluations

Jana-Rebecca Rehse(✉) and Peter Fettke

Institute for Information Systems (IWi) at the German Center for Artificial Intelligence (DFKI GmbH) and Saarland University,
Campus D3 2, Saarbrücken, Germany
{Jana-Rebecca.Rehse,Peter.Fettke}@iwi.dfki.de

Abstract. Given the multitude of new approaches and techniques for process mining, a thorough evaluation of new contributions has become an indispensable part of every publication. In this paper, we present a set of 20 scientifically supported “process mining crimes”, unintentional mistakes that threaten the validity of process discovery evaluations. To determine their prevalence even in high-quality publications, we perform a meta-evaluation of 21 process discovery papers published at the BPM conference. We find that none of these papers is completely crime-free, but the number of crimes and their impact on the evaluations’ validity differs considerably. Based on our list of crimes, we suggest a catalog of 13 process mining guidelines, which may contribute to avoiding process mining crimes in future evaluations. Our objective is to spark an open discussion about the necessity of valid evaluation results among both process mining researchers and practitioners.

Keywords: Process mining · Process discovery · Evaluation
Quality metrics

1 Introduction

Process mining is set out to gain insights into information systems by analyzing their behavior, as recorded in event logs. More specifically, the goal of process discovery is to represent the behavior of the information systems in form of a business process model [1, p. 163ff.]. The quality of process discovery results is often measured in terms of the four dimension fitness (the model’s ability to replay observed behavior), precision (the model’s ability to not allow unobserved behavior), generalization (the model’s ability to explain unobserved behavior), and simplicity (the model’s complexity) [2]. Over the last fifteen to twenty years, a number of process discovery approaches have been proposed [3]. They either address so far unresolved challenges, such as duplicate tasks, or improve the state-of-the-art in terms of result quality as measured by the four dimensions or efficiency, i.e. using less computational resources [4].

M. Weske et al. (Eds.): BPM Forum 2018, LNBIP 329, pp. 3–19, 2018.
https://doi.org/10.1007/978-3-319-98651-7_1

Rehse, J.-R., & Fettke, P. (2018). Process Mining Crimes - A Threat to the Validity of Process Discovery Evaluations. In BPM Forum (pp. 3-19). Springer.

Preventing process mining crimes

UNIVERSITY OF MANNHEIM
Business School

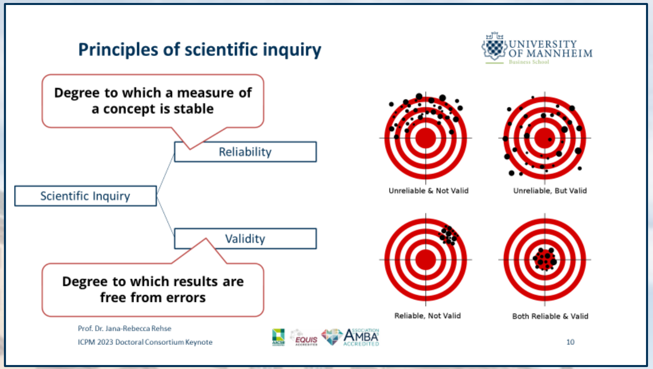
This is hard!!!

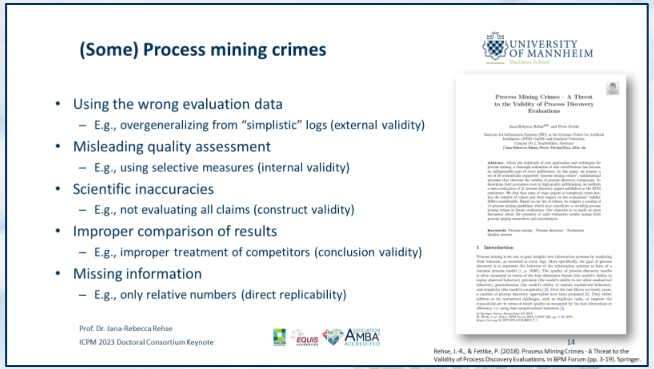
1) Be specific when reporting on your contributions.
2) Explicate assumptions.
3) Choose representative evaluation data and justify this choice.
4) Be aware of the shortcomings of quality measures.
5) Be aware of non-determinism – and don't be afraid of statistics.
6) Make (fair) comparisons to state-of-the-art techniques.
7) Specify your computational set-up, if necessary.
8) Provide the source code and the evaluation data.

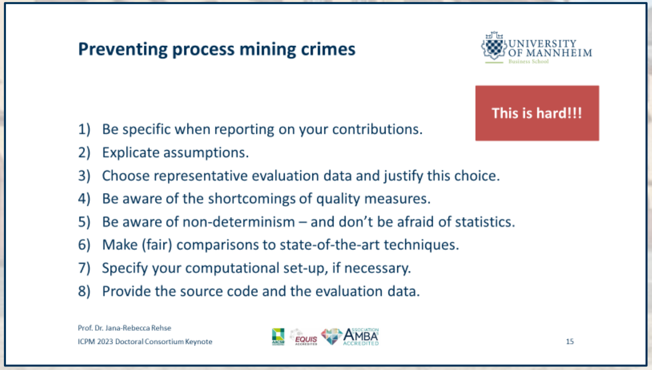
What does this mean for you?

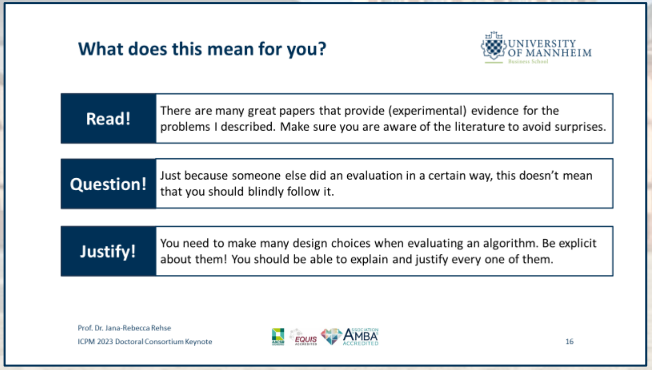
Read!	There are many great papers that provide (experimental) evidence for the problems I described. Make sure you are aware of the literature to avoid surprises.
Question!	Just because someone else did an evaluation in a certain way, this doesn't mean that you should blindly follow it.
Justify!	You need to make many design choices when evaluating an algorithm. Be explicit about them! You should be able to explain and justify every one of them.

“If we understand better what we’re doing, we might be able to do it better”









Prof. Dr.
Jana-Rebecca Rehse

Junior Professor for Management Analytics
University of Mannheim
rehse@uni-mannheim.de