



Human Computer Interaction

Chapter 6: Evaluation

Prof. Dr. Björn Eskofier
Machine Learning and Data Analytics (MaD) Lab
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)
Summer Term 2024



01 Experimental Design

02 Descriptive Statistics

03 Inferential Statistics

04 Heuristic Evaluation

05 Tools for usability evaluation

06 References



Machine Learning and Data Analytics Lab (MaD) in Erlangen



Experimental Design

How do experiments look like?



Machine Learning
Data Analytics

FAU

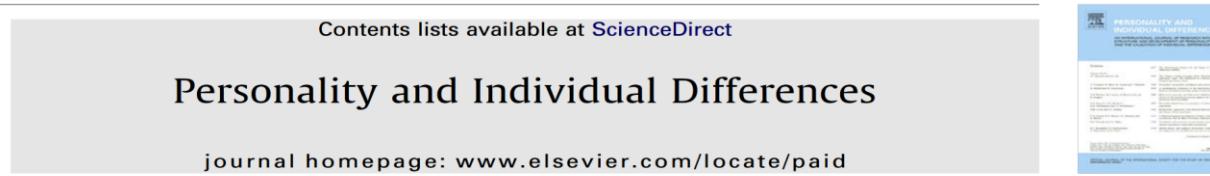


NIHS laboratory, © Nestlé:
<https://www.flickr.com/photos/nestle/8144370795> (CC BY-NC-SA 2.0)



Hypothesis: Negative Impact of Scarring.
How would you design such an experiment?

- **147 female, 76 male** participants rated attractiveness of opposite-sex faces
- Faces were manipulated with photorealistic scarring
- **Scarring** may increase person's attractiveness!



Facial scarring enhances men's attractiveness for short-term relationships
Robert P. Burris ^{a,*}, Hannah M. Rowland ^a, Anthony C. Little ^b



Generalizability of experiments?



M. Hickson, A.L. D'Souza, N. Muthu, T.R. Rogers, S. Want, C. Rajkumar, C.J. Bulpitt.: Use of probiotic Lactobacillus preparation to prevent diarrhoea associated with antibiotics: randomised double blind placebo controlled trial. BMJ. 2007 Jul 14;335(7610):80. Epub 2007 Jun 29.



Scientists want to understand cause and effect

Cause & effect

When metal
is heated
it expands



<https://genagorlin.substack.com/p/time-to-build-the-builders?s=r>

Why we do experiments?



To make **predictions**



Photo by Rich Niewiroski Jr.:
<http://projectrich.com/gallery/> (CC BY 2.5)

The metal in the bridge needs space to expand in hot weather

To test **hypotheses**

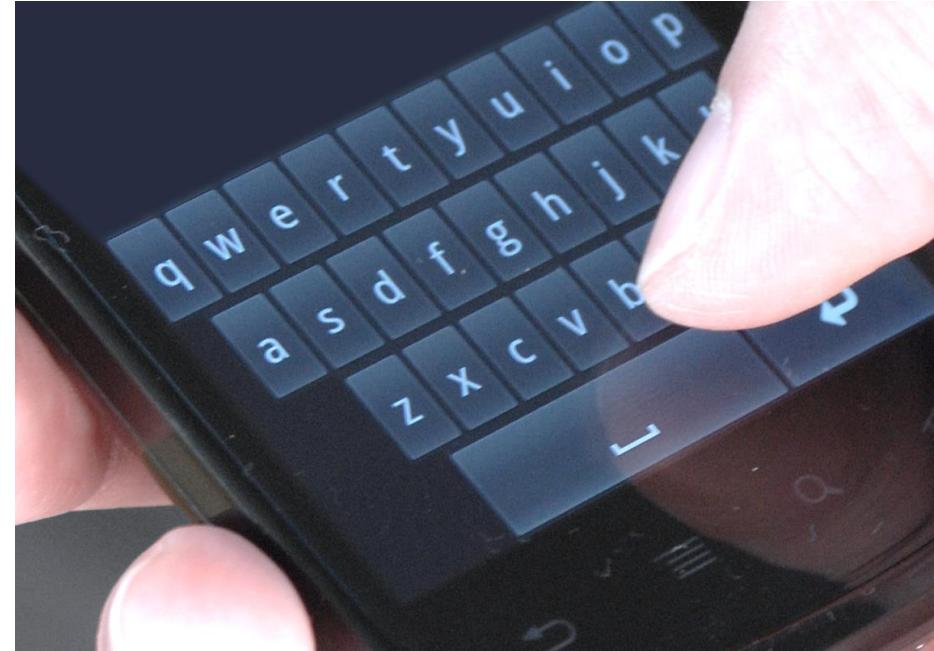


Photo by N. Henze

My Keyboard is faster than yours



Observe users using the user interface and collect data

Why will this not be sufficient?



Example: Keyboard usability



observational studies

The keyboard is easy to use

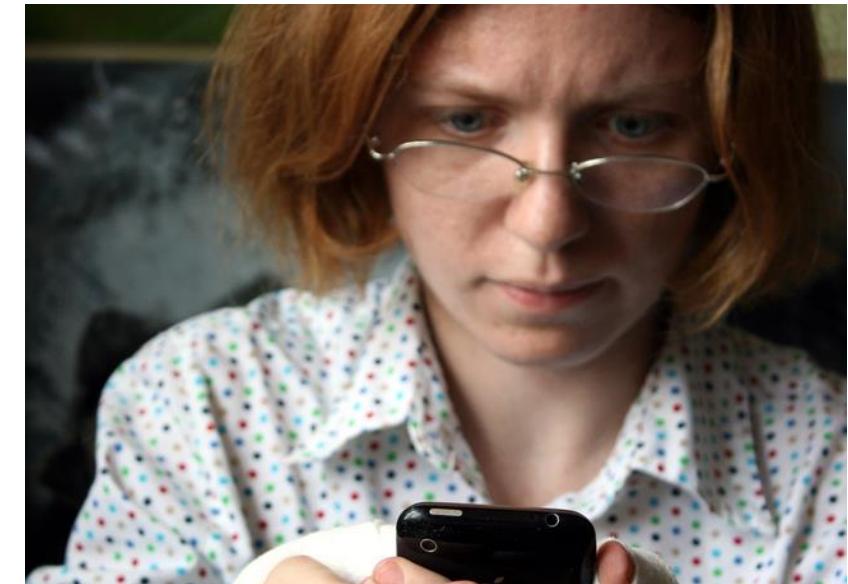
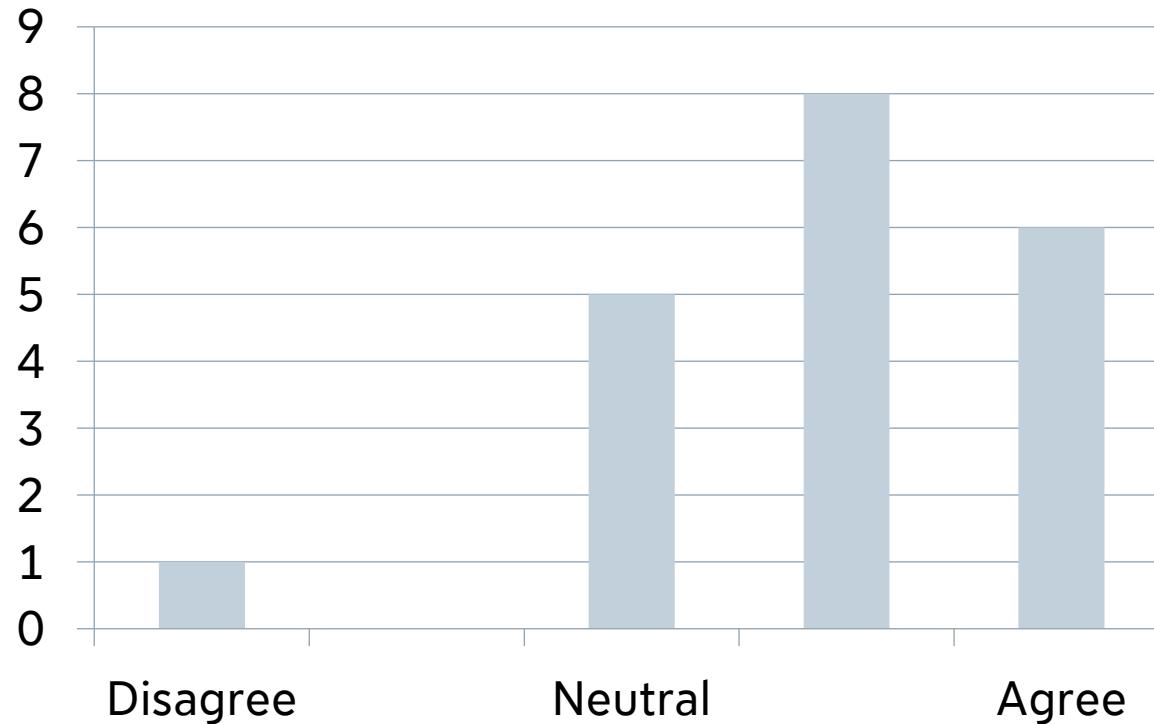


Photo by quinn.anya:
<http://www.flickr.com/photos/quinnanya/4661638173> (CC BY-SA 2.0)



• Isolate cause.

Isolating the cause

Participants rated the keyboard easy to use, because:

- They actually find the keyboard easy to use?
- They want to support you in your research?
- They were overwhelmed by the system's novelty?
- The German football team won the world cup yesterday?
- ...

More observation! 2
easy
riding answer

More observation will not help to find the answer!



Caution vs Correlation.

Storks Deliver Babies ($p = 0.008$)

KEYWORDS:

Teaching;
Correlation;
Significance;
p-values.

Robert Matthews

Aston University, Birmingham, England.
e-mail: rajm@compuserve.com

Summary

This article shows that a highly statistically significant correlation exists between stork populations and human birth rates across Europe. While storks may not deliver babies, unthinking interpretation of correlation and *p*-values can certainly deliver unreliable conclusions.

◆ INTRODUCTION ◆

Introductory statistics textbooks routinely warn of the dangers of confusing correlation with causation, pointing out that while a high correlation coefficient is indicative of (linear) association,

association between storks and the concept of women as bringers of life, and also in the bird's feeding habits, which were once regarded as a search for embryonic life in water (Cooper 1992). The legend lives on to this day, with neonate-bearing storks being a regular feature of greetings cards celebrating births.

Matthews, Robert. "Storks deliver babies ($p = 0.008$)."*Teaching Statistics* 22.2 (2000): 36-38.
<http://www.brixtonhealth.com/storksBabies.pdf>

If I want more babies,
can I move to an area with many storks?

No! Storks do not **cause** babies

Other causes? → other explanations

<http://perfecthealthdiet.com/2012/04/theory-of-the-stork-new-evidence/>

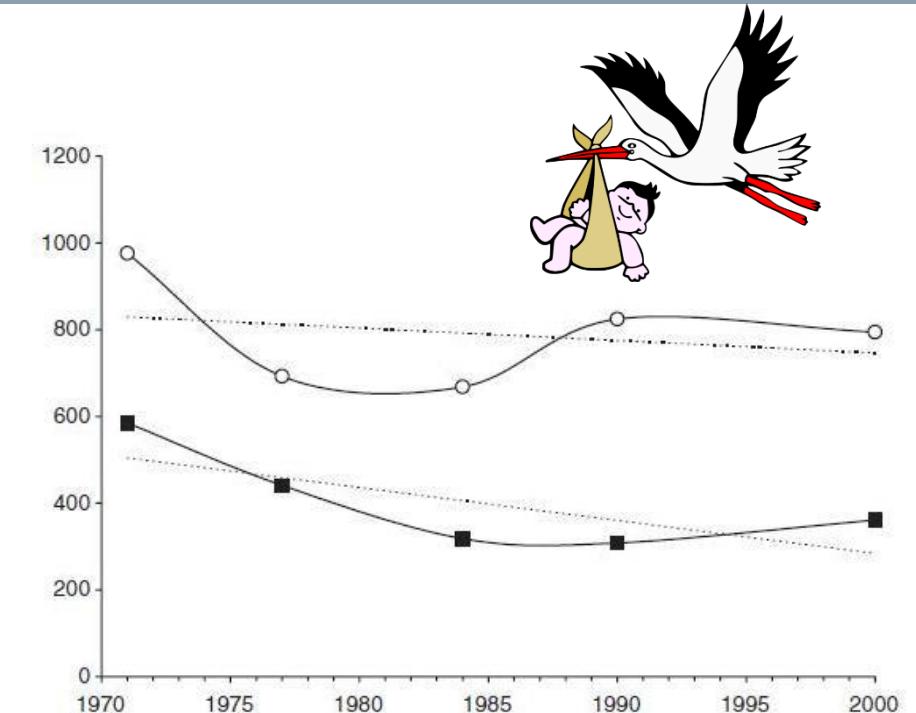


Figure 1. Storks and the birth rate in Lower Saxony, Germany (1971–2000). Open circles show yearly birthrates in hundreds in Lower Saxony. Full squares show numbers pairs of storks in Lower Saxony. Dotted lines represent linear regression trend ($y = mx + b$).



Birthrate and number of storks correlate

Explanation 1: Children cause storks

For example, the crying of babies attract storks

Explanation 2: Storks cause children

For example, the myth is true and storks bring babies

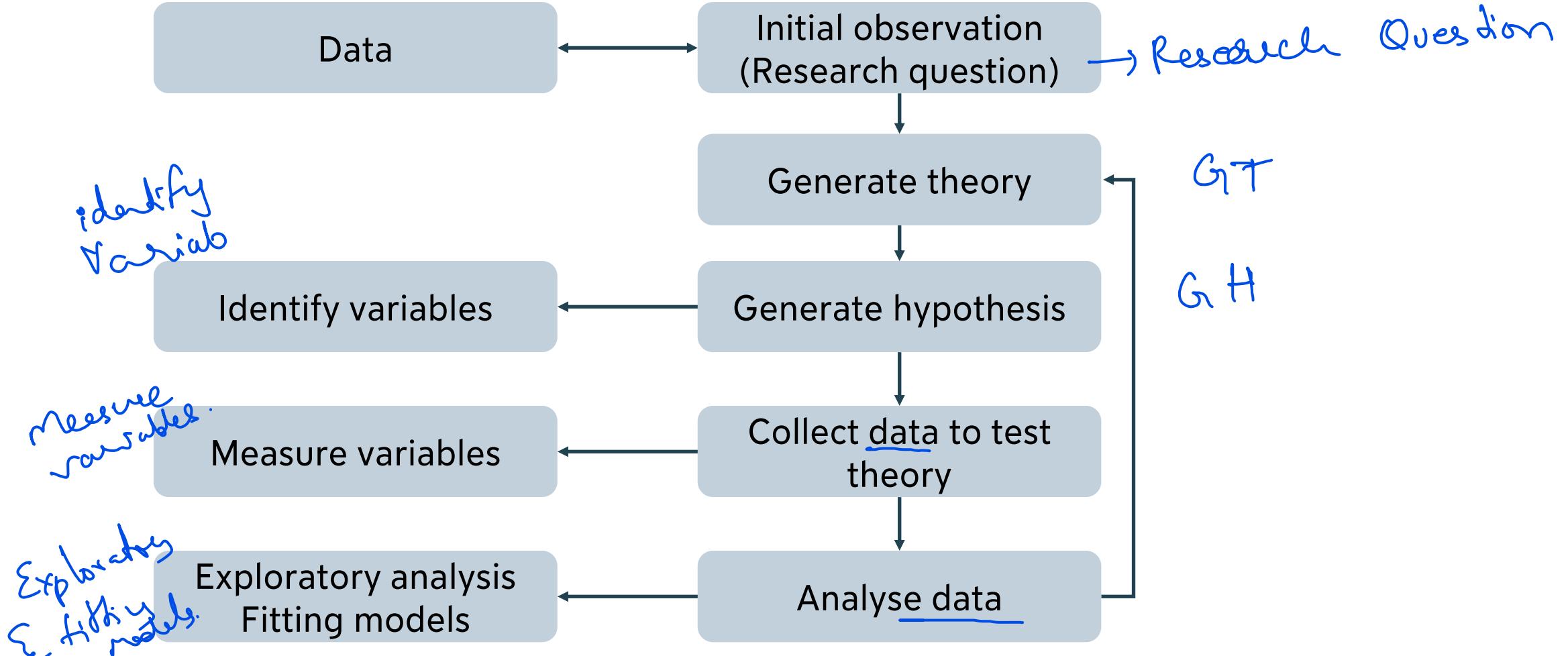
Explanation 3: A third unknown aspect causes both

For example, the village environment is more friendly to storks and families that desire children.

Called Tertium Quid

Unknown explanation with two known -
3rd something.

Research Process





Detect Causal.

Goal: Detection of **causal influences**

Set up a **direct comparison** between **treatments**

Minimize **bias** and **errors**

Main characteristic of experiments and difference to observational studies:

We control the assignment of conditions!

Cause &
Effect.

We can isolate **cause and effect**

Knowing **cause and effect**



**Thank you
for your attention**