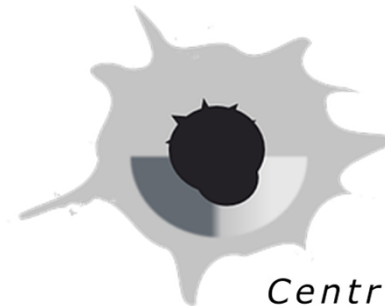
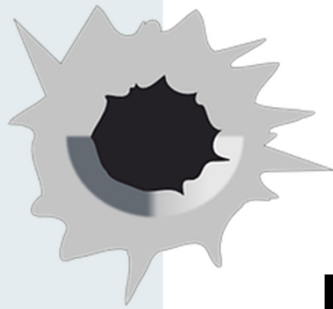


# Kollateralschäden beim Schrotflintenschießen

Big Brother – Big Data – Big Trouble?



**PD Dr. Wolfgang Meyer**

Saarland University  
P.O. Box 151150  
D – 66041 Saarbrücken  
[www.ceval.de](http://www.ceval.de)  
[w.meyer@mx.uni-saarland.de](mailto:w.meyer@mx.uni-saarland.de)

*Centrum für Evaluation*



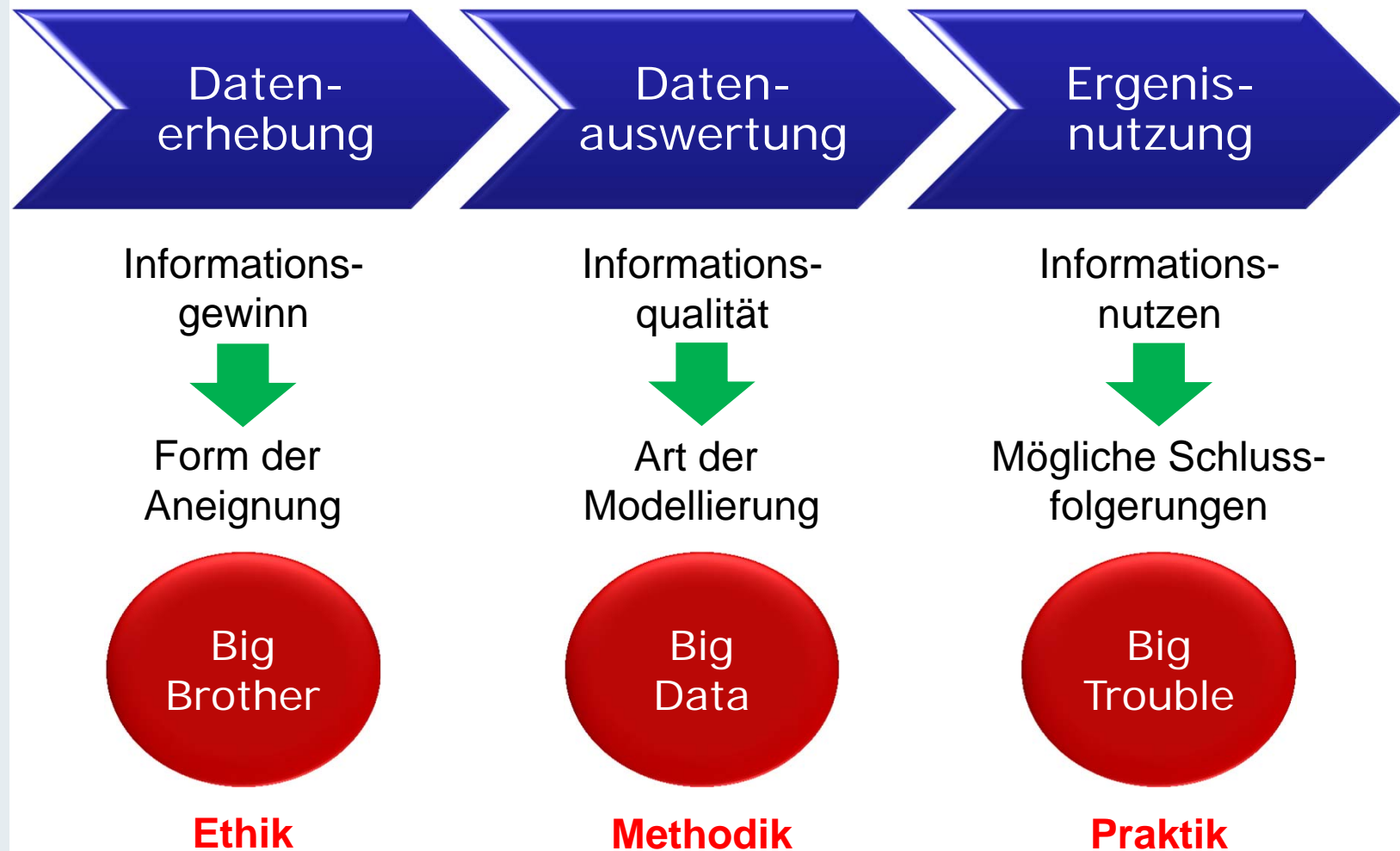
*Center for Evaluation*

## Zentrale Fragestellungen

- ✓ Sollten Evaluationen Big Data nutzen?
- ✓ Welche Kriterien sollten für diese Entscheidung herangezogen werden?
- ✓ Welche Risiken und Nebenwirkungen sind dabei zu beachten?

- ✓ **Prolog**
- ✓ **Big Brother**  
Datenerhebung
- ✓ **Big Data**  
Datenauswertung
- ✓ **Big Trouble**  
Datennutzung
- ✓ **Epilog**

## Was zu bewerten ist...



# Kriterien zur Bewertung der Datenerhebung

- ✓ Erfolgt der Informationstransfer freiwillig und ethisch korrekt?
- ✓ Lassen sich Störfaktoren in der Erhebungssituation eliminieren oder kontrollieren?

# Kriterien zur Bewertung der Datenanalyse

- ✓ Sind die Daten valide und reliabel?
- ✓ Ist die statistische Modellierung adäquat zur Approximation, Erklärung und/oder Vorhersage?
- ✓ Lassen sich die statistischen Modelle überprüfen und empirisch testen?

## Kriterien zur Bewertung der Datennutzung

- ✓ Erfolgen die Bewertungen anhand klar definierbarer Evaluationskriterien?
- ✓ Lassen sich auf der Grundlage der Ergebnisse objektive, neutrale und nachvollziehbare Entscheidungen vornehmen?
- ✓ Lassen sich Risiken und Nebenwirkungen der Entscheidungen abschätzen und kontrollieren?

Die **ethische Komponente** – wem  
gehören die Daten und wer darf sie  
für was verwenden?



## „Hackerangriffe durch Malware möglich“

*(Alex Gibney, Dokumentarfilm „ZeroDays“)*

Software mit dem Decknamen "Olympic Games": Sie "saugt" nicht nur aus beliebig-vielen Quellen für Big Data, sondern hat die Macht, derartige Datenmengen substantziell zu verändern und dadurch in Sekunden ganze Staaten lahmzulegen oder individuelle Existenzen zu vernichten.

## Manipulationsmöglichkeiten

**Standardisierung der Erhebungstechnik ermöglicht die bewusste Manipulierung**

## „Dank Psychometrie gewann Trump die Wahl“

*(Hannes Grasegger und Mikael Krogerus, Artikel „Ich habe nur gezeigt, dass es die Bombe gibt“)*

Jeder Charakterzug eines Menschen lässt sich anhand von fünf Persönlichkeitsdimensionen messen und gut vorhersagen: Offenheit, Gewissenhaftigkeit, Extraversion, Verträglichkeit und Neurotizismus (Michael Kosinski).

## Missbrauch durch Prognosequalität

„Regelmässigkeit des Verhaltens ermöglicht die bewusste Manipulierung“

## „Fehlende Abgrenzung privater und öffentlicher Daten“

*(Jan Voß, Politikwissenschaftler an der Uni Kiel)*

Machtgefälle zwischen Staat und Wirtschaft gegenüber den Bürgern durch die unterschiedlichen Möglichkeiten des Datenzugangs und der Datenmanipulation.

## Entmachtung der Bürger

„Ungleichverteilung der Zugangsmöglichkeiten zu Daten“

## Zusammenfassung:

- ✓ *Informationsträger* können bewusst Daten manipulieren
- ✓ *Informationsnutzer* können bewusst Daten manipulieren
- ✓ *Manipulationen* lassen sich schwer nachweisen
- ✓ Bewertung der Datenqualität und *Steuerung des Erhebungsvorgangs* ist kaum möglich
- ✓ Bestehende *Machtunterschiede* werden verstärkt –Big Data ist undemokratisch

Die **methodische Komponente** –  
werden Modelle durch mehr Daten  
tatsächlich immer besser?

**„Mehr Daten führt zu mehr Datenmüll“**

*(Hans Ulrich Buhl, Professor für Finanz- und Informationsmanagement Uni Augsburg)*

Mit der Menge der Daten steigen die Probleme, die relevanten Daten herauszufiltern

## **Herausforderung für das Datenmanagement**

**„Mehr Daten heißt ... weder ... automatisch bessere Daten, ... bessere Entscheidungen oder zufriedene(ere) Kunden. “**

## „Korrelationen statt Kausalität“

*(Gary Marcus und Ernest Davis,  
Psychologieprofessoren New York University,  
„Eight – no Nine – Problems with Big Data“, NYT)*

Big Data erschließt nicht den Sinn der  
Analyseergebnisse

## Sinnlose Modellierung

**„If you look one hundred times for  
correlations you will find by chance five  
even if there is no connection“**

## „Algorithm Dynamics influence quality“

*(David Lazer, Ryan Kennedy, Gary King und Alessandro Vespignani, Harvard und Northeastern University Bosten, „The Parabel of Google Flu: Traps in Big Data Analysis“ (Science)*

Massenpanik verändert die Prognosen

## Feedback Effekte bei Prognosen

**„Most Big Data are not the output of instruments designed to produce valid and reliable data“**



## Zusammenfassung:

- ✓ Große Datenmengen müssen *adäquat gemanagt* und verarbeitet werden
- ✓ Große Datenmengen ignorieren *Unterschiede in der Datenqualität*
- ✓ Große Datenmengen produzieren *„Rauschen“* welches Effekte überdeckt
- ✓ Große Datenmengen *verstellen den Blick auf die Kausalität* (sind immer „signifikant“)
- ✓ Große Datenmengen können *Zyklen produzieren* („Self-full filling prophecies“)

Die **Wirkungskomponente** –  
worin bestehen Risiken und  
Nebenwirkungen?

**„es gibt keine neutralen Algorithmen“**  
(*Corinna Bath, Informatikerin TU München*)

Technik ist immer von sozialen  
Umständen beeinflusst

## **Pseudo-Objektivität**

**„Menschen, die Algorithmen nutzen,  
hinterfragen diese in der Regel nicht,  
sondern nehmen die gelieferten  
Ergebnisse als wahr und richtig hin.“**

## „Diskriminierung durch Algorithmen“

*(Michael Kneissler, „Wenn Big Data falsch liegt“, Beitrag im Wissenschaftsmagazin PM)*

Bewertungen erfolgen mechanisch  
in einer Blackbox anhand von  
unbekannten Variablen

## Automatisierte Evaluationen

**„Algorithmen sind sexistischer als unsere  
heutige Gesellschaft.“**

**„Maschinen verstehen keine Zwischentöne“**

*(Christian Rätsch, Unternehmensberater)*

Daten produzierte Scheinrealitäten  
fördern defensives, konservatives  
Entscheiden

## **Mechanisierung von Verhalten**

**„Datenanalysen sind der Versuch,  
Schlüsse aus der Vergangenheit  
für die Zukunft zu ziehen“**

**„Menschen werden auf Datenspuren reduziert“**

*(Christian Scholz, BWL-Professor der  
Universität des Saarlandes)*

Auf Theorien, Hypothesenbildung und jeglichen  
Anflug von wissenschaftlicher Vorgehensweise  
wird verzichtet..

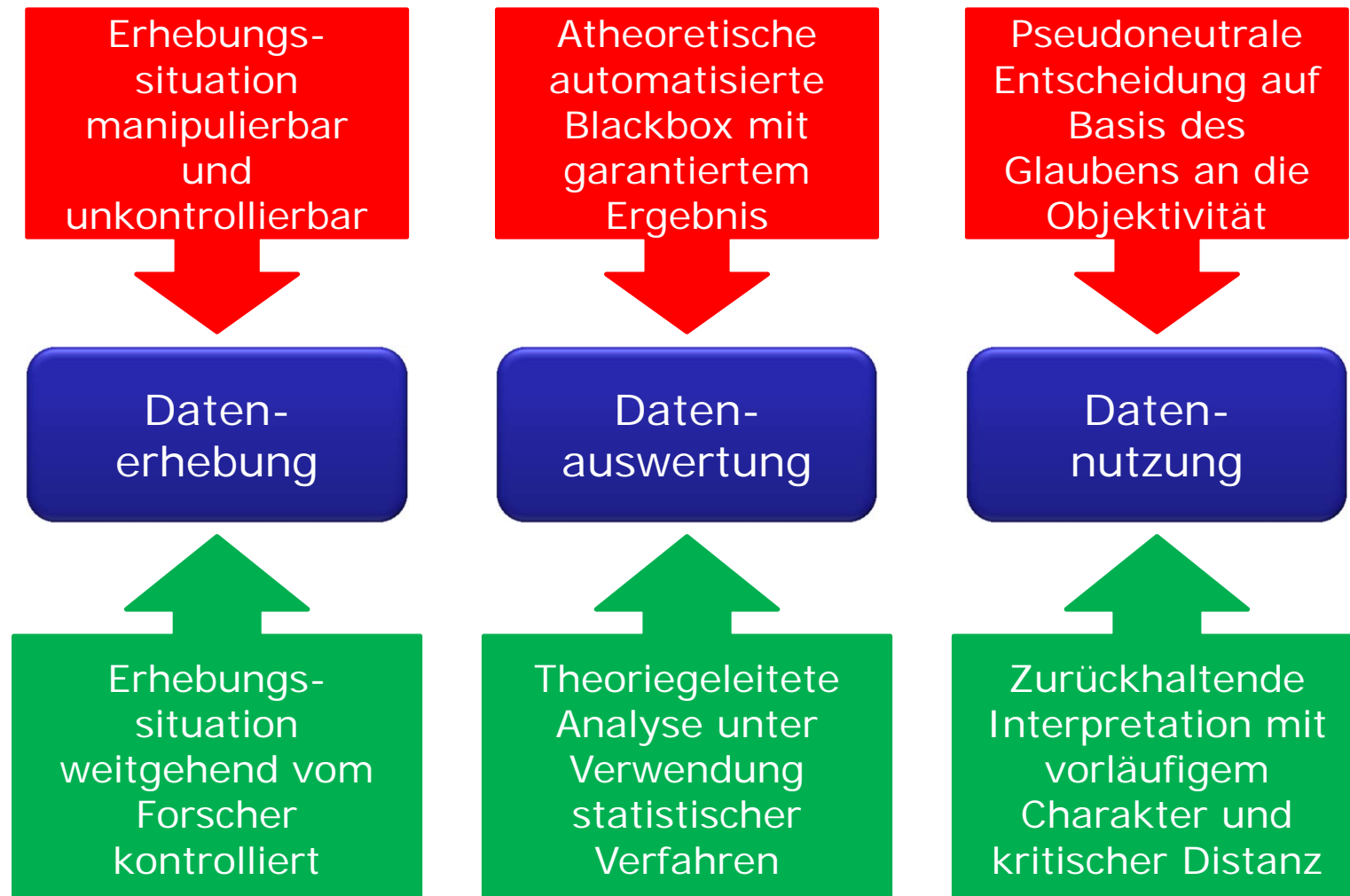
## **Atheoretisches Handeln**

**„Man nimmt irgendwelche Daten und freut  
sich ohne weiteres Nachdenken über  
scheinbare Zusammenhänge, die dann sofort  
("pragmatisch") zu Handlungen führen “**

## Zusammenfassung:

- ✓ Die Evaluation erfolgt automatisch und scheinbar objektiv („personenunabhängig“)
- ✓ Vorurteile und Diskriminierungen werden datengestützt konserviert
- ✓ Unangenehme Entscheidungen werden der „objektiven“ Maschine überlassen
- ✓ Die Interpretationsmacht liegt beim Programmierer („Black Box“)
- ✓ Die Entscheidungsgrundlagen sind nicht überprüfbar, Fehlentscheidungen werden nicht korrigiert

## Big Data



## Wissenschaftsmodell



## Zentrale Antworten

- ✓ Sollten Evaluationen Big Data nutzen?

**Nur unter besonderen Ausnahmebedingungen**

- ✓ Welche Kriterien sollten für diese Entscheidung herangezogen werden?

**Freiwilligkeit und Transparenz der Beteiligung**  
**Weitgehende Kontrolle der Erhebungssituation**  
**Offenlegung der Algorithmen und Berechnung**  
**Überprüfbarkeit der Ergebnisse**  
**Reflektierte Nutzung als eine (nicht die!)**  
**Datenquelle**

## Zentrale Antworten

- ✓ Welche Risiken und Nebenwirkungen sind dabei zu beachten?

**Selbstbestimmungsrecht der Informationsträger**

**Manipulationsmöglichkeiten**

**Eingeschränkte Erklärungskraft**

**Alternativer und zusätzlicher Methodeneinsatz**