C!S

Centre for Informatics & Society

TU WIEN **!**

FAKULTÄT
FÜR !NFORMATIK

Faculty of Informatics

Lectures Series: Aspects of Digital Transformation
Monday, December 10th, 2018

Data – Interests – Ontologies: How business models distort science

Klaus Kornwachs

Universität Ulm, acatech München, Berlin,
Büro für Kultur und Technik, Argenbühl

---

Data – Interests – Ontologies: How business models distort science

1. New possibilities to find new insights?

2. Science method and business methods

3. To put the right question remains an art

4. Do numerical approaches substitute science?

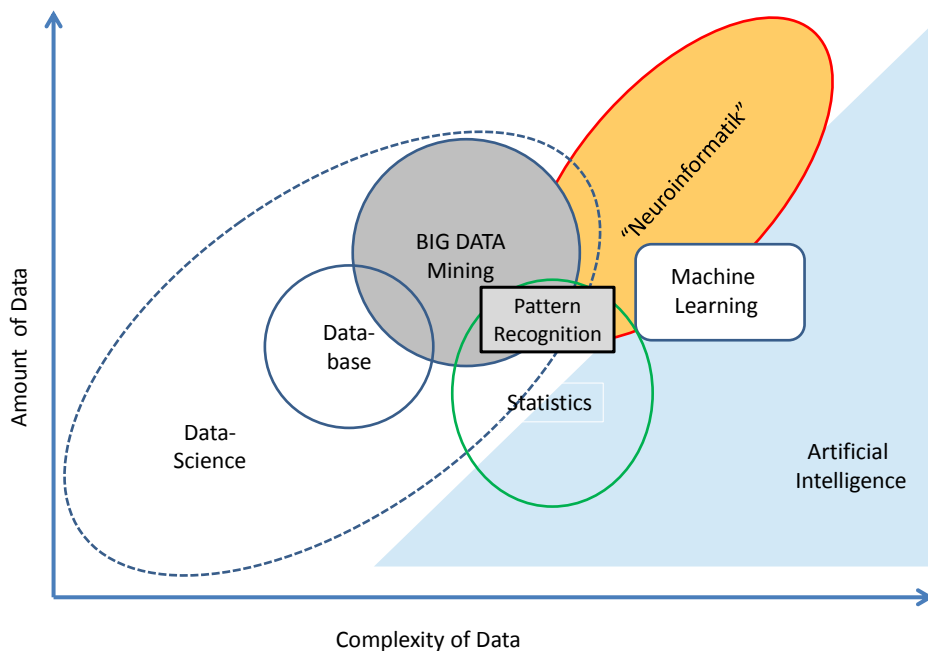5. Two examples: Geomarketing and HR-Management

6. Responsibility

The new possibilities of Big Data and Artificial Intelligence Technologies allow finding unexpected structures and relations in data sets, generated in different fields and contexts.
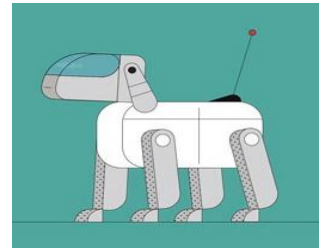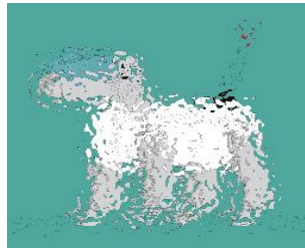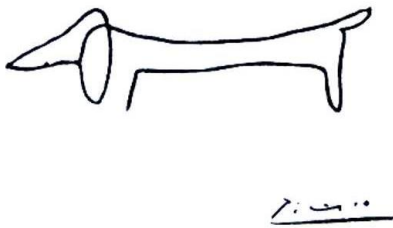
Portfolio of Basics for Big Data (acc. to Mainzer 2014)

Tab. 9, K. 4

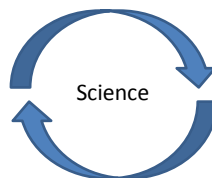|	Descriptive Analytics	Predictive Analytics	Prescriptive Analytics
Qusetion	States and dynamics	What, when und why who	What should be done and why
Methods	Reports Dashboards Scorecards Data Warehouse	Data Mining Text Mining Web Media Mining Extrapolation /Fitting	Optimization Simulation Decision Modeling Expert Systems
Resuts	Business reports and Options	Forecasts	Options to decide Preferences

From Data to Decision

BIG DATA ?

| Result | 1 | 5 | 2 | 3 | 3 | 4 | 5 | 6 | 2 | 4 | 1 | 6 | 2 | 5 | 3 | 2 | 1 | 6 | 3 | 5 | 2 | 4 |
| Throw | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |

Data

Analytics
mask, arrangement
calculation
visualization

Throwing: Data collection from process

111 22222 3333 444 5555 666

Science

Zahl 6   3
Zahl 5   4
Zahl 4   3
Zahl 3   4
Zahl 2   5
Zahl 1   3

Interest: Are the dices biased?

Information

Decision
A More Throws
B Remove the dice, they may be faked

Knowledge incl. Pre-Knowledge
A The data are not sufficient to decide
B Number 2 is really biased

Cognition:
Understanding information
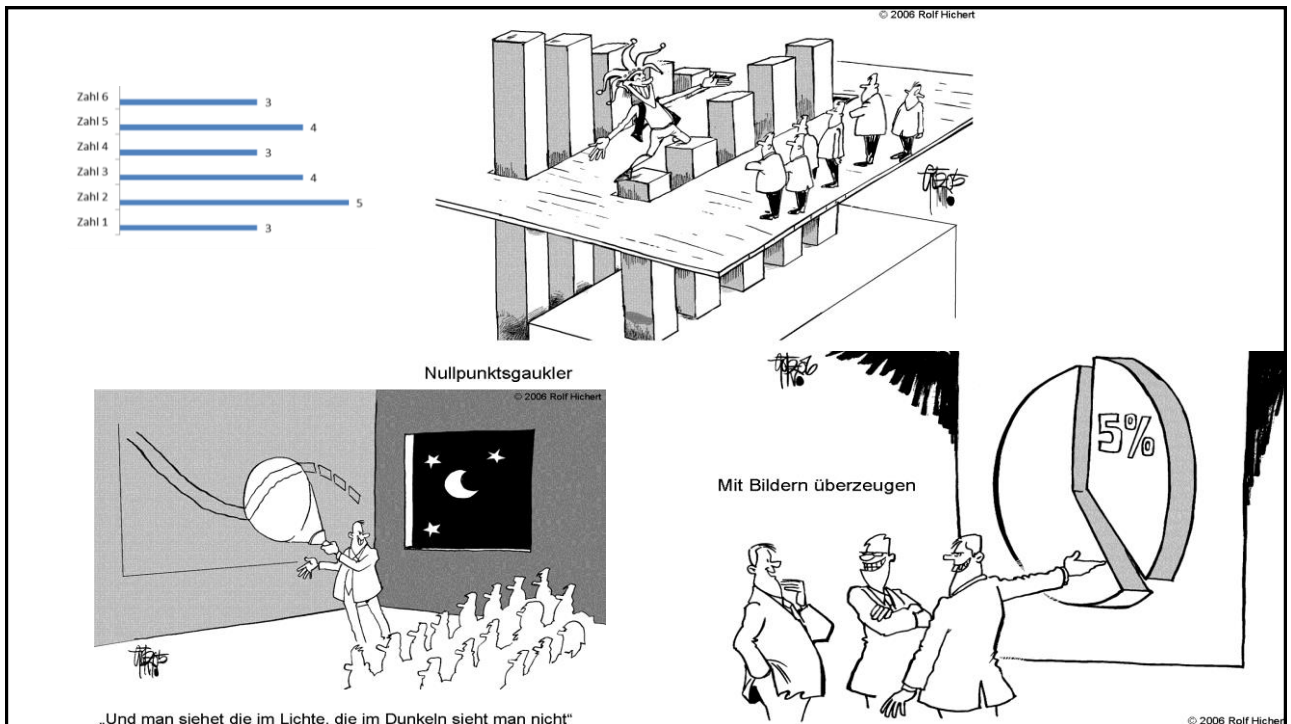
3

Nullpunktsgaukler

Mit Bildern überzeugen

„Und man siehet die im Lichte, die im Dunkeln sieht man nicht"

Data – Interests – Ontologies: How business models distort science

1. New possibilities to find new insights?

2. **Science method and business methods**

3. To put the right question remains an art

4. Do numerical approaches substitute science?

5. Two examples: Geomarketing and HR-Management

6. Responsibility

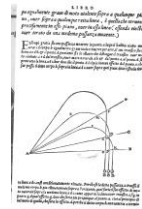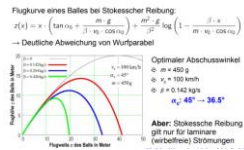## Scientific Method: How to win knowledge on order to act (I)

$v_0 = 150$ m/sec
$\alpha = 45°$
$h = 1,80$ m

https://www.rpguides.de/darkest-dungeon/komplettloesung/kanone-der-briganten-brigand-pounder.htm

Preparation of initial- and boundary conditions

ceteris paribus

"Nature" Device Machine

cannot be influenced → Observation B free form interference

Measurement

can be influenced with A

Knowledge of initial- and boundary conditions

Tables Diagramms Equations Groups Symmetries

Calculation Computer Solutions

I-O Behavior: Data A, Data B

Predictions Cal of B = Function of (A) due to theory A → B

Pragmatic Syllogism (no theorem)

If A → B and	$\mathcal{B}!$	Then try	B per $\mathcal{A}$
Theoretical Knowledge	**Goal**	**Pragmatic Knowledge**	**Technological Rule**
Causal relation Experimentally confirmed by comparision between observed I-O behavior and prediction by theory	State $\mathcal{B}$ should be achieved $!$	Causal relation has to do with real ontology	Rules are effective or not

A → B is part of real world

Select Action $\mathcal{A}$ to put A into practice

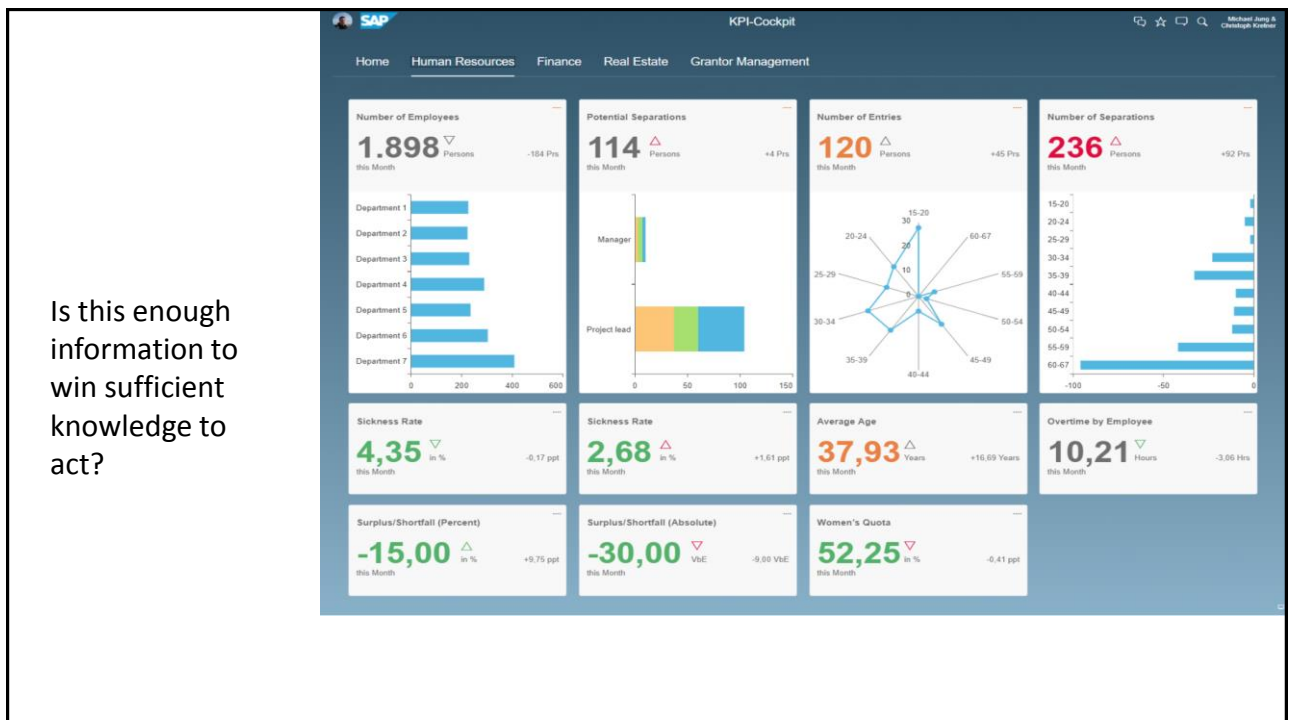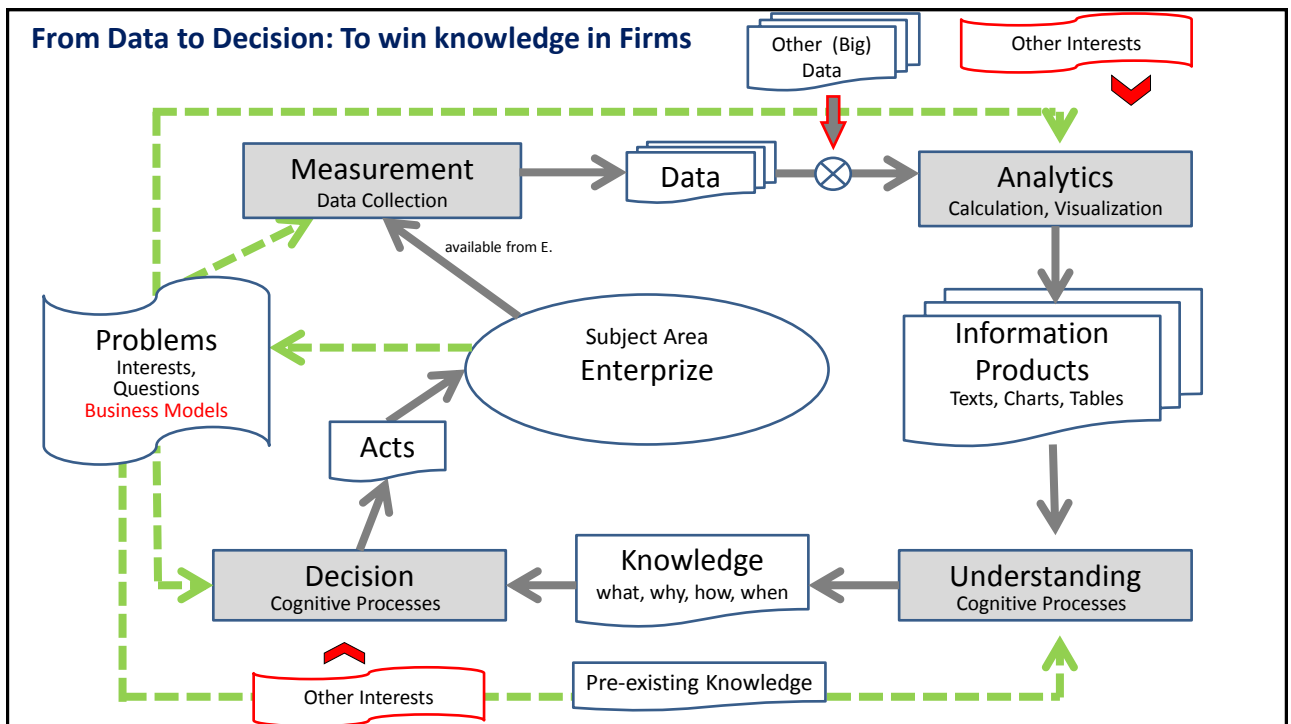$\mathcal{A}$ is feasable

It is possible to act sucessfully only by knowing the rule

Data (A) and Data (B)

Cal (B) = Function (A)
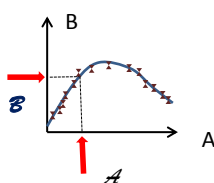
Are data analytics (BIG DATA) sufficient to establish a certain probability for the effectivity of a technological rule to act?

From Data to Decision: To win knowledge in Firms

Is this enough information to win sufficient knowledge to act?

Available information (Files, Books, Charts, Interactive Means) is not yet knowledge!
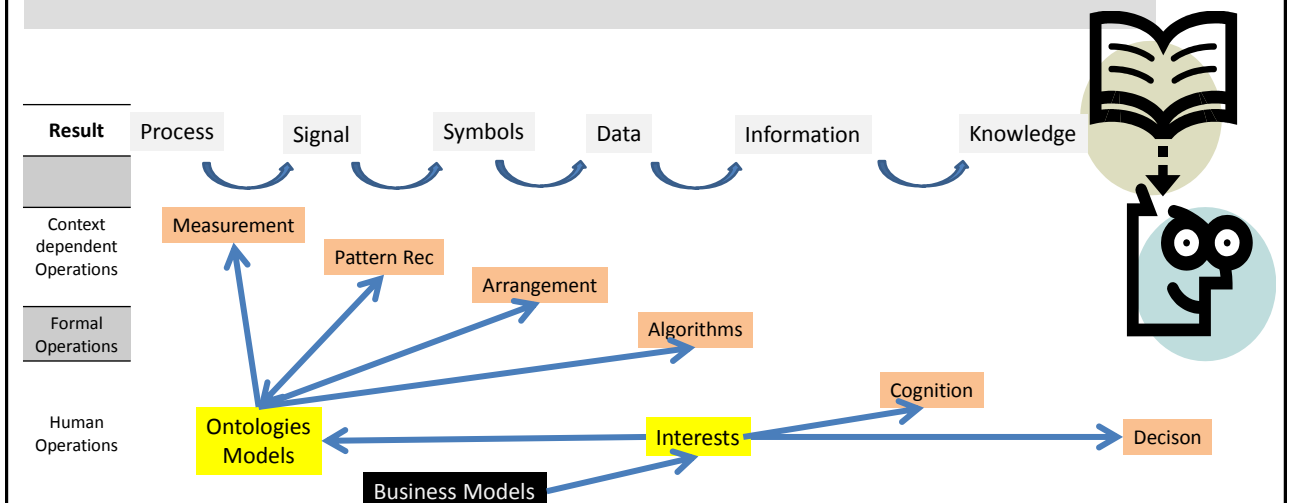
| Result | Process | Signal | Symbols | Data | Information | Knowledge |

Context dependent Operations

Measurement
Pattern Rec
Arrangement
Algorithms

Formal Operations

Human Operations

Ontologies Models
Interests
Cognition
Decison
Business Models

Data – Interests – Ontologies: How business models distort science

1. New possibilities to find new insights?

2. Science method and business methods

3. **To put the right question remains an art**

4. Do numerical approaches substitute science?

5. Two examples: Geomarketing and HR-Management

6. Responsibility

Abb. 19, K. 3
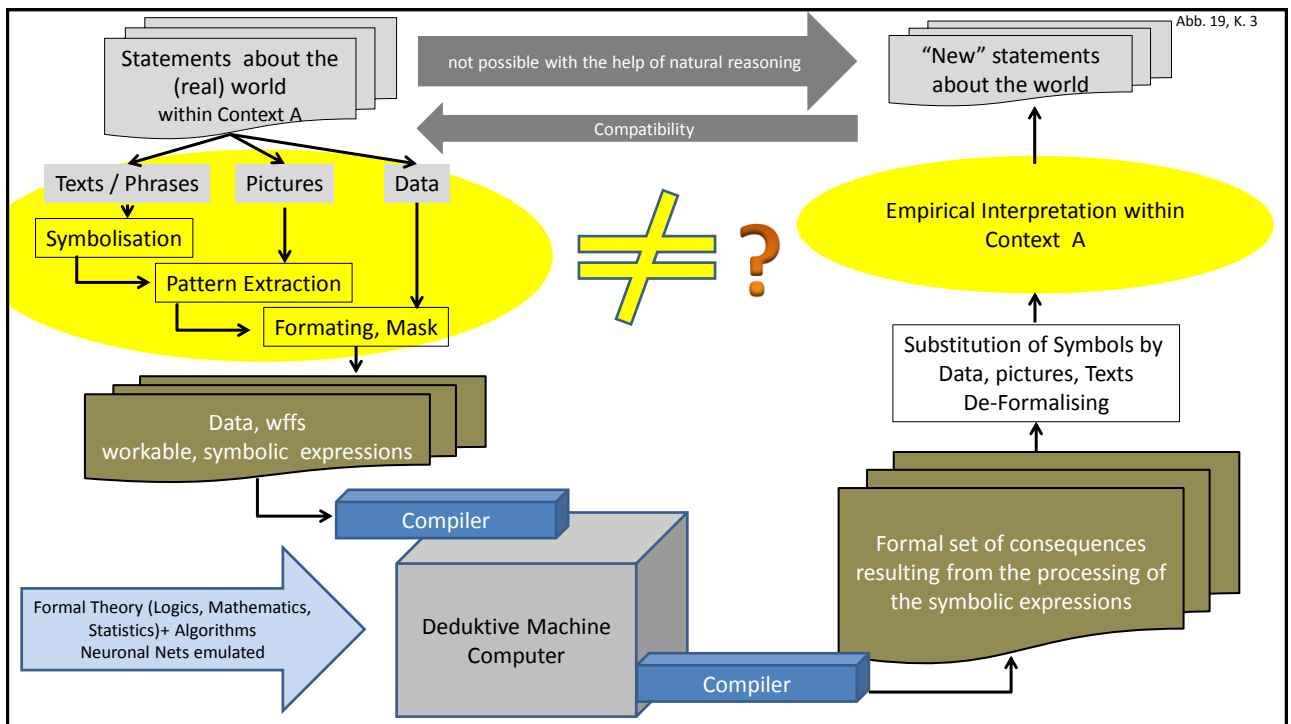
Information may produce knowledge.

decontentualized information, i.e. information not adequately interpreted may lead to a wrong, i.e. misleading knowledge

If knowledge is one of the most important presuppositions of the ability to act, wrong knowledge may lead to wrong acts.

Reasoning:

Deductiv: $((A \rightarrow B) \wedge A) \rightarrow B$

Abductiv: $((A \rightarrow B) \wedge B) \rightarrow A$ logically wrong! only fuzzified: $\rightarrow_{\mu}$
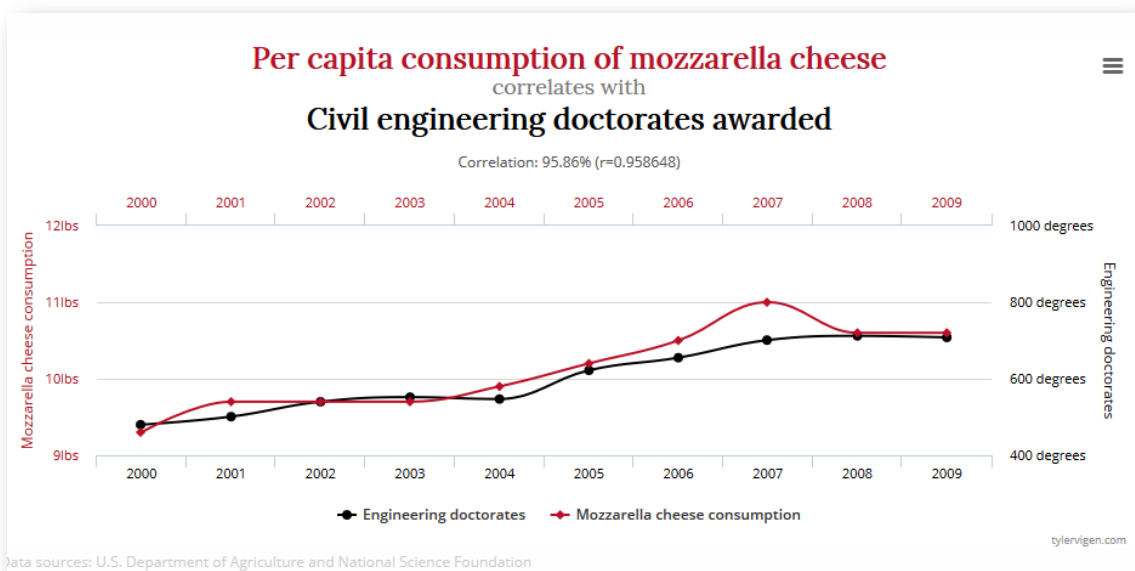
Inductiv: (very often $A \wedge B$), then $(A \rightarrow B)$ with $p \Rightarrow$ empirical science

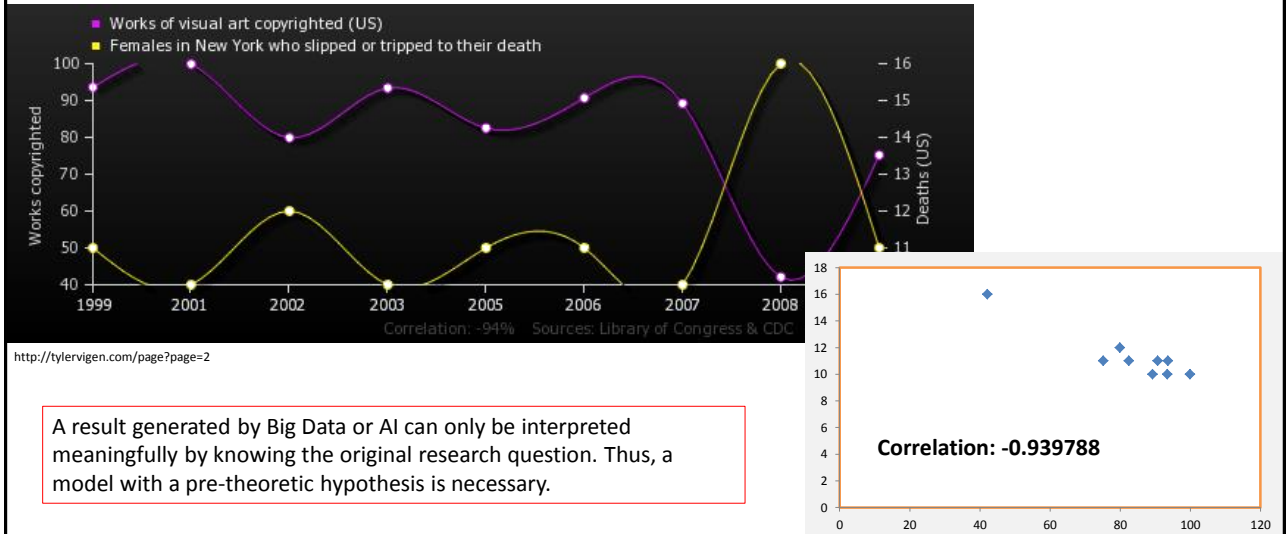**Context change leads to change of interpretation**

Source: Frey (1984, 2000).

Meaningful Question? Hypothesis with low probability in terms of causal effect Findings by chance? Statistical Correlation is no proof

Works of visual art copyrighted (US)
is inversely correlates with
Females in New York who slipped or tripped to their death

http://tylervigen.com/page?page=2

A result generated by Big Data or AI can only be interpreted meaningfully by knowing the original research question. Thus, a model with a pre-theoretic hypothesis is necessary.

Correlation: -0.939788

Abductive reasoning is used constitutively in Big Data Methods. Therefore it is not an adequate method to ground jugments in the realm of e.g. Human Resource Management (People Analytics) and related fields.

Knowledge won by Big Data Analytics **information** is not yet reliable and should only used for the finding of hypotheses. Such hypotheses should be tested iteratively by more data, experiments, observation, inquiries.

Due to cost and capacity reasons such additional tests are used to be omitted in business context.

To take responsiblity for success and reliabilty of the use of BIG DATA results remains with the user of such systems. This responsiblity cannot be delegated.

Data – Interests – Ontologies: How business models distort science

1. New possibilities to find new insights?

2. Science method and business methods

3. To put the right question remains an art

4. Do numerical approaches substitute science?

5. Two examples: Geomarketing and HR-Management

6. Responsibility

All data, available by social networks, scientific data, company data and collected within the net, due to inquiries and data bases form a

"*massive corpus as a laboratory of the human condition*"

The storage of Petabytes of data within offered clouds

"*... forces us to view data mathematically first and establish a context for it later*"

(Chris Anderson: The End of Theory: The Data Deluge makes the Scientific Method Obsolete. https://wired.com(2008)/06/pb-theory/.

"The point I was making -- and I don't remember the exact words -- was that if the model is going to be wrong anyway, why not see if you can get the computer to quickly learn a model from the data, rather than have a human laboriously derive a model from a lot of thought."
Googles Research Director Peter Norvig in reply to Anderson (2008).

"The new availability of huge amounts of data, along with the statistical tools to crunch these numbers, offers a whole new way of understanding the world. Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all."

only as a first step …

presumably not!

Anderson (2008).

"We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot."

Anderson (2008)

meaningful?

"The idea is that businesses collect massive sets of data that may be homogeneous or automatically collected. Decision-makers need access to smaller, more specific pieces of data from those large sets. They use data mining to uncover the pieces of information that will inform leadership and help chart the course for a business."

(Techopedia 2015).

"I spend half of my advertising budget for free, and I'd like to know which ones …"

(Rockefeller)

?????

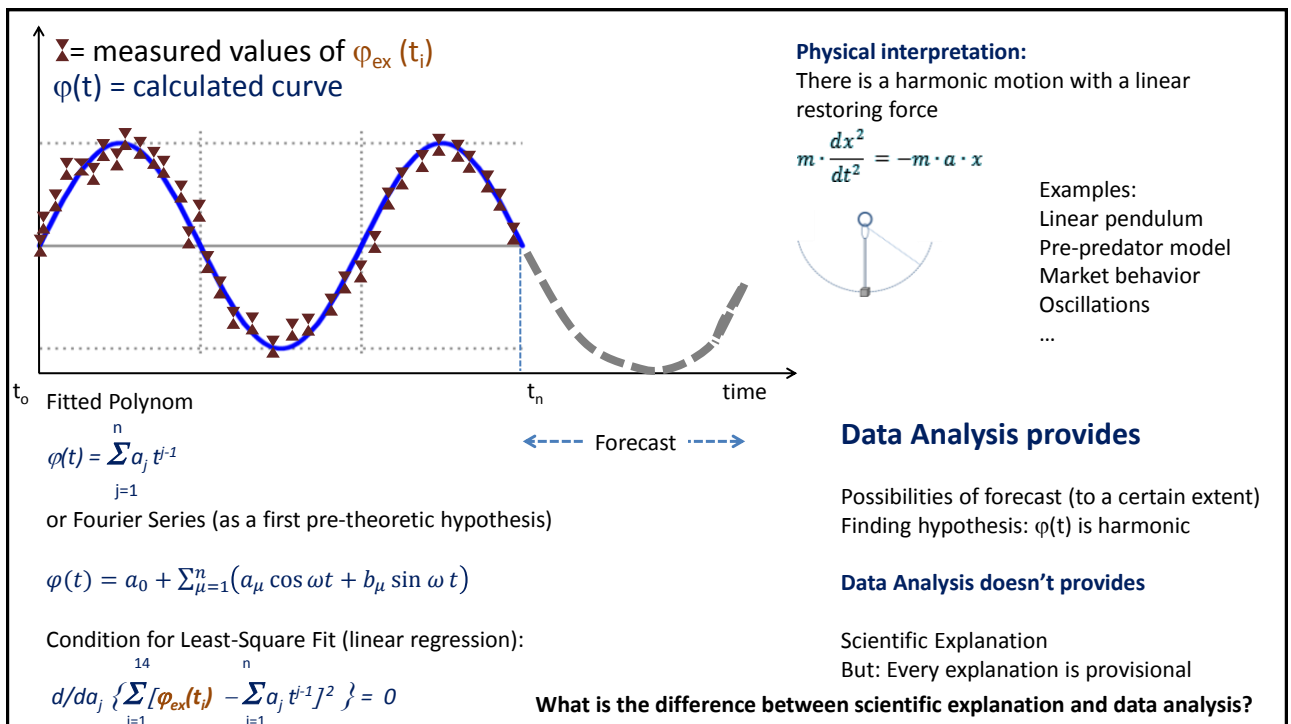*"Since Aristotle, we have fought to understand the causes behind everything. But this ideology is fading. In the age of big data, we can crunch an incomprehensible amount of information, providing us with invaluable insights about the what rather than the why."*

(Mayer-Schonberger & Cukier, 2013)

refrain from sience?

What is the difference between scientific explanation and data analysis?

= measured values of $\varphi_{ex}(t_i)$
$\varphi(t)$ = calculated curve

Physical interpretation:
There is a harmonic motion with a linear restoring force

$$m \cdot \frac{dx^2}{dt^2} = -m \cdot a \cdot x$$

Examples:
Linear pendulum
Pre-predator model
Market behavior
Oscillations
…

t_o Fitted Polynom

$$\varphi(t) = \sum_{j=1}^{n} a_j\, t^{j-1}$$

t_n time

‹---- Forecast ----›

or Fourier Series (as a first pre-theoretic hypothesis)

$$\varphi(t) = a_0 + \sum_{\mu=1}^{n}(a_\mu \cos \omega t + b_\mu \sin \omega t)$$

Condition for Least-Square Fit (linear regression):

$$d/da_j \left\{ \sum_{i=1}^{14} [\varphi_{ex}(t_i) - \sum_{j=1}^{n} a_j\, t^{j-1}]^2 \right\} = 0$$

Data Analysis provides

Possibilities of forecast (to a certain extent)
Finding hypothesis: $\varphi(t)$ is harmonic

Data Analysis doesn't provides

Scientific Explanation
But: Every explanation is provisional

What is the difference between scientific explanation and data analysis?

Even if we had found an elegant technical solution of **finding a needle in a haystack**, we would need to have a model presentation of the needle.

With appropriate filters and new combinations and mask arrangements it is **always possible to find** something unplanned.

However, the question is: What do these structures and patterns tell us, or better: What do we believe **that they should tell us** something at all?

Without theory it is possible to provide **multiple data interpretations**, including everything and nothing.

Nevertheless: Without theory, we have to start with data analysis and **then** proceed further **iteratively.**

The criticism of BIG DATA should remain fair
Hildago (2014)

✓ The use of BIG DATA in scientific and technology context: Enormous progress

Data Rich Web Sites,
Information visualization,
Machine learning,
Machine translation,
Data base technologies,
Social media sites,
Social networks
Computer vision,
Facebook algorithms with face matching,
Proliferation of computational design,
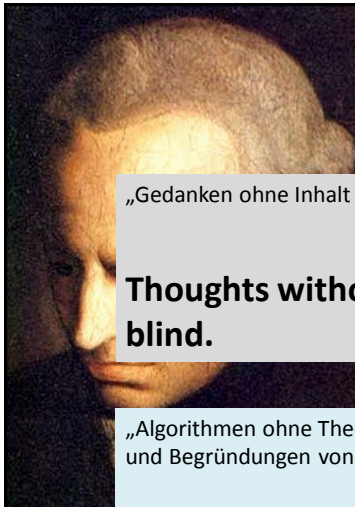Data journalism,
Artistic expressions (animation)

✓ Critiques of claims no serious data scientist has ever made.

✓ Confusion of the limitation of prototype with fatal flaws.

✓ The iterative character of development is overseen

✓ Wrong predictions

Nevertheless:

„Gedanken ohne Inhalt sind leer, Anschauungen ohne Begriffe sind blind."
Kant, Immanuel: Kritik der Reinen Vernunft. B 75, A 51. Vgl. Kant (1956), S. 95.

Thoughts without perception are empty, intuitions without concepts are blind.

„Algorithmen ohne Theorie und Gesetze sind blind! Korrelationen und Datenmuster ersetzen keine Erklärungen und Begründungen von Ursachen. Daher brauchen wir Urteilskraft."

Algorithms without Theory and Laws are blind! Correlation and data patterns cannot substitute explanations and groundings of causal relations. Therefore we need power of judgement.

Mainzer (2016).

Data – Interests – Ontologies: How business models distort science

1. New possibilities to find new insights?

2. Science method and business methods

3. To put the right question remains an art

4. Do numerical approaches substitute science?

5. **Two examples: Geomarketing and HR-Management**

6. Responsibility

Geomarketing (State of the art: Rauner 2007)

Data from:
House inspections, surveys, mail order and debtor information, address traders, government agencies, companies inquiries, orders by internet, Consumer Scan

Mean value of at least 5 households

Results for Kupfergraben 6, 10117 Berlin

- ✓ built before 1900,
- ✓ six households, construction satisfactory,
- ✓ no garden available,
- ✓ no foreigners,
- ✓ affinity for loyalty cards: medium,
- ✓ affinity for private health insurance: medium,
- ✓ residents: disinterested financial type,
- ✓ classic landline users,
- ✓ hardly internet power users,
- ✓ dominating age: 51 to 60 years ,
- ✓ Diabetes and osteoarthritis above average,
- ✓ fitness below average,
- ✓ a lot of Audi, Mercedes and BMW, little Volkswagen.

Interested in such data:
Energy supplier,
Credits Institutes, drective bureaus,
baknks Postbank, IT-provider, Firms like
Rossmann, Vorwerk, Arcor,
ADAC, Pharmaceutical companiers
Supermarkets, Assurance companies,
Food Stores Marketing, Advertising
Companies etc.

"Living in the wrong place, ... can lead to worse conditions for obtaining a loan"
(Peter Schaar, cited in Rauner, 2006, p. 38)

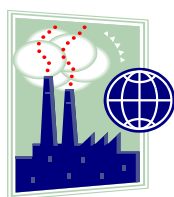**Geomarketing** (State of the art: Rauner 2007)

Interests and Business Models:

❑ **Marketing:** centralized data management for arbitrarily structured data, generated by different sources like Web analytics, Social Media, advertising, campaigning

❑ **Organization:** real-time systems for inventory management, warehousing, dispatch handling

❑ **Process optimization:** optimization of individual processes within complex production chains

❑ **Trade:** sales forecast, trend analysis, retargeting, predictive buying

❑ **Internet of things:** connected cars, smart traic control, smart city Media analysis: difusion of contents, Web analytics and others

❑ **E-Commerce, on-line Shops, frequencies of use "Industry 4.0":** data analysis of real-time production processes and applications for control purposes

(*um & HP, 2014)

The conclusions from this "discovery" were straight forward, but eventually in conflict with the ideal of democracy, equality, and fairness.

Abb. 20, K.4

2000 Applications /Day ?

Personalwesen –
Human Ressource Management

Personal Recruiting (PR)

Profiling

Capacity Management

Administration (salary, times, places)

Leadership (PF)

Qualification on the Job

Image care

Labour Market

Market of Orders
(Crowd-Working)

Work
Tasks
Projects

Skills Supply

Skills Demand

Abb.23, K. 4

Abb. 22, Kap.4

Mathematic Procedures
Statistics, Database, Analytics, Big Data,
developed for further purpuses

HR Administration
Payroll Managemnt

regular EXCEL

Applications

Data of Labour Market

Simple Matching

HR-Development

Formalising,

Standardising
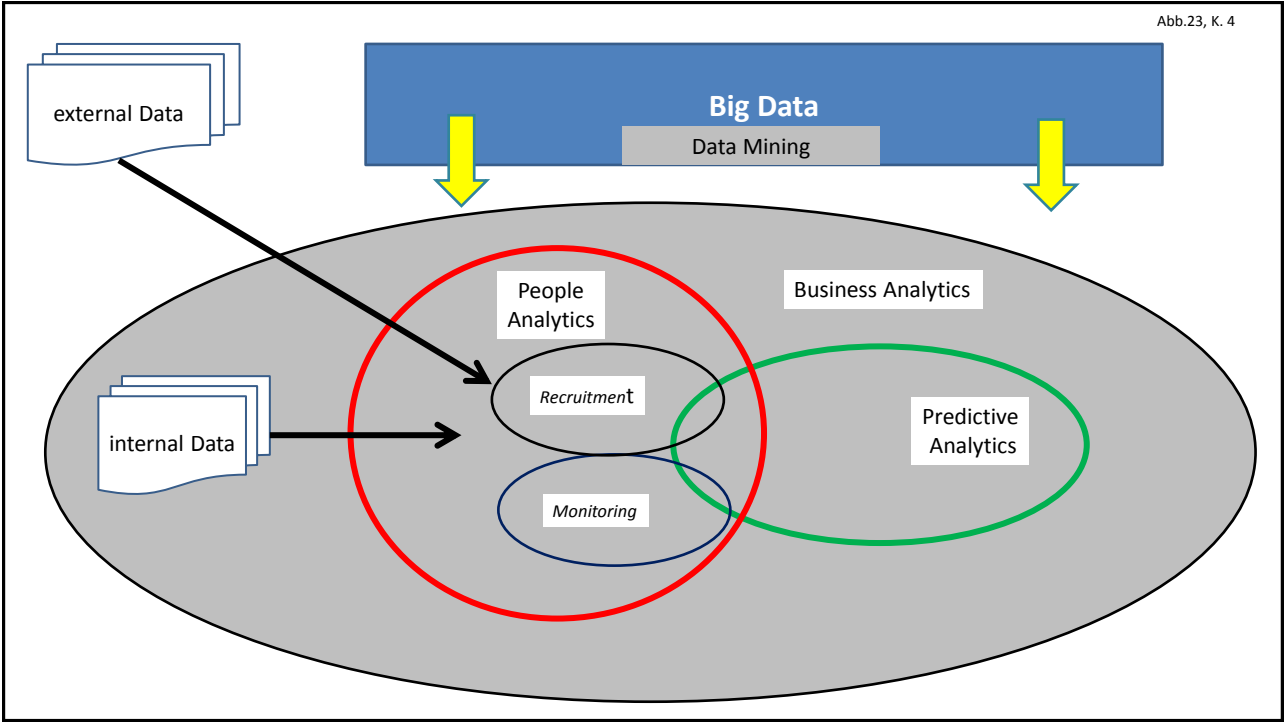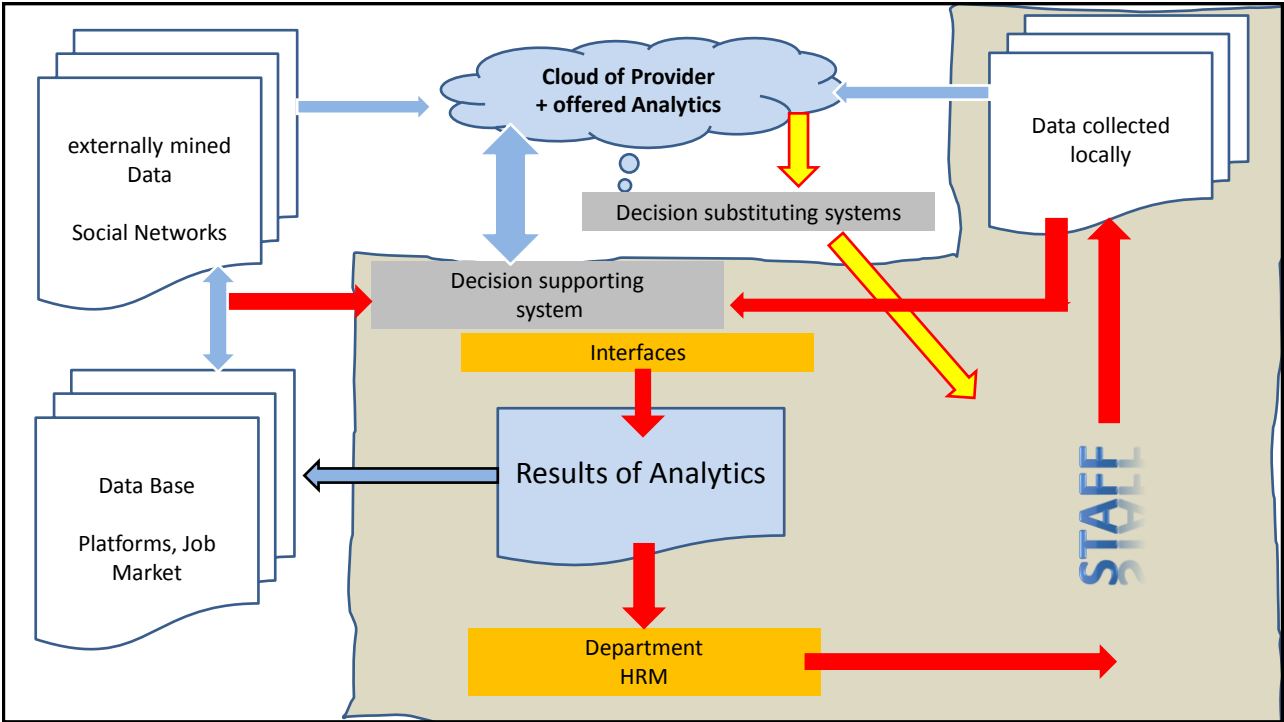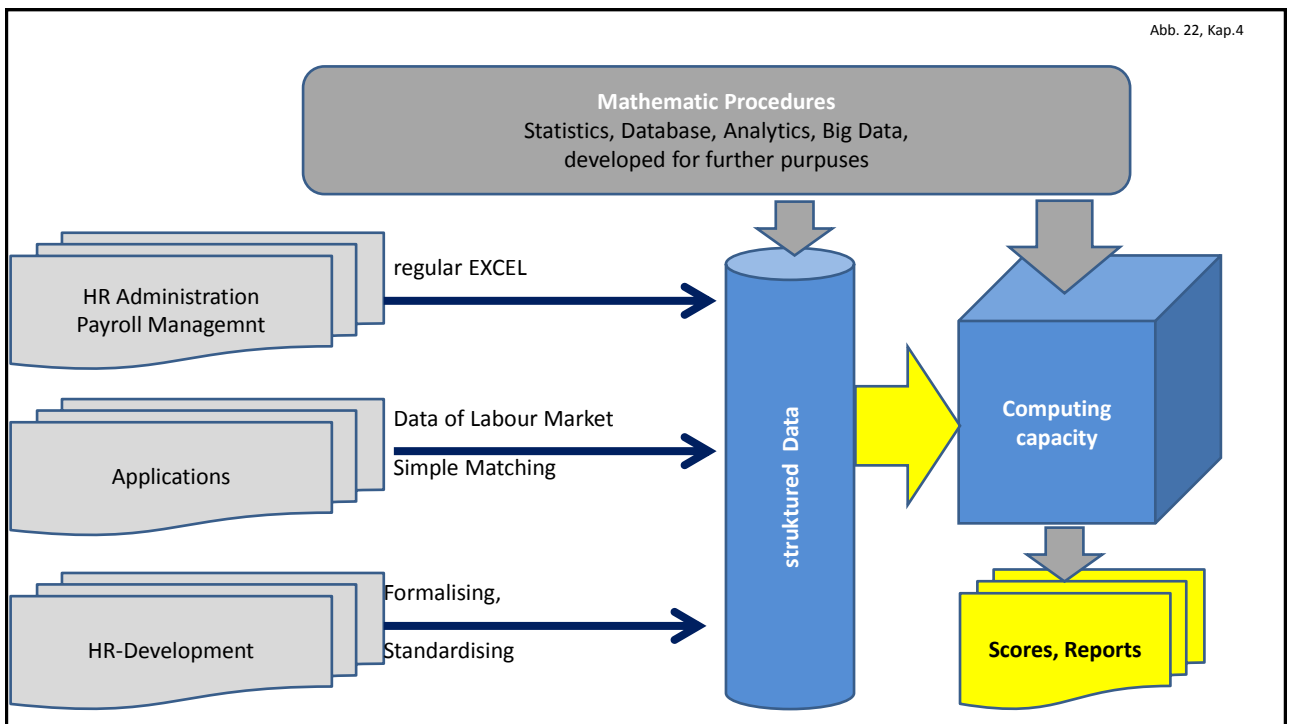
struktured Data

Computing capacity

Scores, Reports

Hoped-for effects and reasons of an introduction of People Analytics:

- Better accuracy of available data (incl. external data),
- Acceleration of HR- decisions (recruitment, monitoring),
- Reliable decisions,
- Reduction of time costs and capacities ,
- Improved strategic planning for recruitment,
- Better measurement (capture) of perfomance und mood of the staff,
- Lower the rate of fluctuation.

Predictive Factors for termination of employment
(Dorio et al 2014)

"Professional mobility"	Influence
Relations with a job broker	+
Contentment with job	-
Efficiency of the responsible manager	-
Oportunities	-
Number of obtained promotions	-
Competitiveness in pay and incentives	-
Competitiveness of Benefits (Incentives, Privileges)	-
Received wage increases in%	-
Appreciation	-
Workplace for sure	-
Layoffs in the organization	+
Outsider positions in the organization	+
Safety at work	-
Time spent with teleworking	+
work stress	+
additonally	
Sex	+ -
Employment	
Level of Hierarchy	+
High Potential Status	+
Branch of Industry	~
Country	~

Steps from voluntary to coercive:

- Wearables are completely voluntary.

- The transfer of data of wearables is completely voluntary. The analysis is free.

- The insurance / HR-Management offers a premium waiver /benefits for voluntary distribution.

- The insurance only lowers the premium if analysis of the data suggests a low risk of disease, i.e. if the insured is preventively effective. /The HR-Management offers only adavantages if …

- The insurance raises the premium massively for all who do not provide their data.

- The insurance insured nobody more without readiness to provide data. / Wearables are prerequisite for an appointment

Decision-making replaced with the help of people analytics prevents necessary participation

The use of People Analytics in Decision-Substituting Mode will cause conceivably to a reinforcement and prolongation of inheritance effects and will stigmatize applicants by past data.

Conveivably it will promote discrimination.

Responsibility gap: Who is responsible for the decision?

The question of the reversibility: It arises when using so-called autonomous systems or fully automated systems. It has not been clarified under which criteria they can be shut off by the user.

Kornwachs (2018)

Data – Interests – Ontologies: How business models distort science

1. New possibilities to find new insights?

2. Science method and business methods

3. To put the right question remains an art

4. Do numerical approaches substitute science?

5. Two examples: Geomarketing and HR-Management

6. **Responsibility**

The question of big data usage comes to a head with the question of a decision-making system that could make personnel decisions on its own. A distinction must be made between smaller companies, large companies and platforms offering services in the area of human resource management. Despite possible technical and economic benefits, the concerns outweigh.

Although the diffusion of big data technology in the field of human resources has not progressed as far as some interested parties suggest, it is recommended to apply already strict standards for the use of such systems

From a theoretical point of view, it turns out that the results of personal analytics are also not very reliable. Moreover their handling can lead to ethically dilemmatic situations, possibly violating legal regulations such as informational self-determination, rights of participation and personal rights.

In general, one can argue that fully-ruling and autonomous systems should not be used in areas where personal security and integrity, core values, human rights, and definable responsibilities (including liability) play a role.

References

Anderson, Ch.: The End of Theory. The Data Deluge Makes the Scientific Method Obsolete. WIRED Magazine June 23rd 2008, http://www.wired.com/2008/06/pb-theory/ .

Frey, Siegfried: Die Macht des Bildes - Der Einfluss der nonverbalen Kommunikation auf Kultur und Politik". Huber, Bern 2000

Frey, Siegfried: Die nonverbale Kommunikation. Stiftungs-Reihe / SEL-Stiftung für Technische und Wirtschaftliche Kommunikationsforschung im Stifterverband für die Deutsche Wissenschaft, Bd. 1. Stuttgart 1984

Hidalgo, Cesar, (MIT Media Lab): Saving Big Data from Big Mouths – Those who would condemn big data ought to try making something Scientific American April 29th, 2014

Kornwachs, K.: *Das Prinzip der Bedingungserhaltung.* Lit, Münster, London 2000

Kornwachs, K.: Ebenen der Orientierung. Zur Analytik des normativen Hintergrundes. In: Hubig, Chr. (Hrsg.): *Ethische Ingenieursverantwortung – Handlungsspielräume und Perspektiven der Kodifizierung*. Sigma, Berlin 2003, S. 31-49 und S. 105-130

Mainzer, Klaus: Zur Veränderung des Theoriebegriffs im Zeitalter von Big Data und effizienten Algorithmen. In: Berliner Debatte Initial 27 (2016), Heft 4, S. 22-34

Mayer-Schönberger, V.; Cukier, K. (2013). Big Data: A Revolution That Will Transform How We Live, Work and Think. London: John Murray.

Norvig, Peter: All we want are the facts, ma'm (2008). http://norvig.com/fact-check.html

Rauner, Max: Die Merkels von nebenan. ZEIT Wissen, 4 (2006), 36-41

Techopedia (2015). What is the difference between big data and data mining? http://www.techopedia.com/7/29678/technology-trends/what-is-the-difference-between-big-data-and-data-mining.

Vigen, Tyler: Spurious Correlations. Hachette Books, 2015. Examples in: www.tylervigen.com/spurious-correlations

_____

Prof. Dr. phil.habil. Dipl. Phys.
Klaus Kornwachs
Former Chair for Philosophy of Technology, BTU Cottbus
Humboldt Study Center for Humanities, University of Ulm
Honorary Professor, China Intelligent Urbanization Co-Creation Center at
Tongji University, Shanghai

Office for Culture and Technology
Strickers Höhe 22
D-88260 Argenbühl
Germany
Phone: ++49-7566-2498, Fax: -1891
Mobil: ++49-175-4122079
klaus.kornwachs@uni-ulm.de
klaus@kornwachs.de
Home Page
www.kornwachs.de