Human-Centered AI
Research Seminar

Module 2: The Fundamentals: Theory of Science

Andreas Holzinger
Human-Centered AI (Holzinger Group)
Institute for Interactive Systems and Data Science, TU Graz
Institute for Medical Informatics, Statistics & Documentation, Medical University Graz
and
Explainable AI-Lab, Alberta Machine Intelligence Institute, Edmonton, Canada

@aholzin #KandinskyPatterns

Course Homepage: https://human-centered.ai/hcai-research-seminar-2020

This is the version for printing and reading. The lecture version is didactically different.

- In Module 2 we discuss questions including “What is science?”, “*Why* contributing to the international scientific community?” and some methodological issues, whereas
- in Module 3 we discuss questions including “*How to* contribute to the international scientific community?” and learn the basic mechanics of science, the “know-how”,
- Of course always from our human-centered AI and machine learning perspective …

01 Why should we contribute to the international scientific community?

- Science can provide explanations *)
- Science can make predictions
- *) these are the questions of “why” something is the cause (cf. Judea Pearl – see course 706.315)
- Why is a person doing this – anthropological explanations formed ancient philosophy – understanding and explaining nature …
- An explanation is a type of insight (sensemaking)
- When’s a good explanation good? Obviously, when we **“feel”***) that something is satisfactorily explained – when we do not have any more questions – we understand it!

*) please watch the video shown on slide 7 by Richard Feynman

- 1) Causal explanation: If something causes X, it also explains X
- 2) Functional explanation: X has a reliable function f(X), thus X is explained
- 3) Purposeful explanation: X was wanted by Y
- 4) Pragmatic explanation: The explanation is adapted to the type of answer the questioner wants to hear!

Explanation "why" – needs a ground truth

https://www.youtube.com/watch?v=MO0r930Sn_8

This was one motivation to develop the Kandinsky Patterns, our "Swiss-Knife" for the study of explainable AI: https://human-centered.ai/project/kandinsky-patterns

Science is humans greatest collective endeavour

UNESCO

United Nations Educational, Scientific and Cultural Organization

"Building peace in the minds of men and women"

IN BRIEF | WHAT WE DO | WHERE WE WORK | PARTNERS | JOIN US | RESOURCES

Home › Science for Society

Science for Society

Science is the greatest collective endeavor. It contributes to ensuring a longer and healthier life, monitors our health, provides medicine to cure our diseases, alleviates aches and pains, helps us to provide water for our basic needs – including our food, provides energy and makes life more fun, including sports, music, entertainment and the latest communication technology. Last but not least, it nourishes our spirit.

Read more

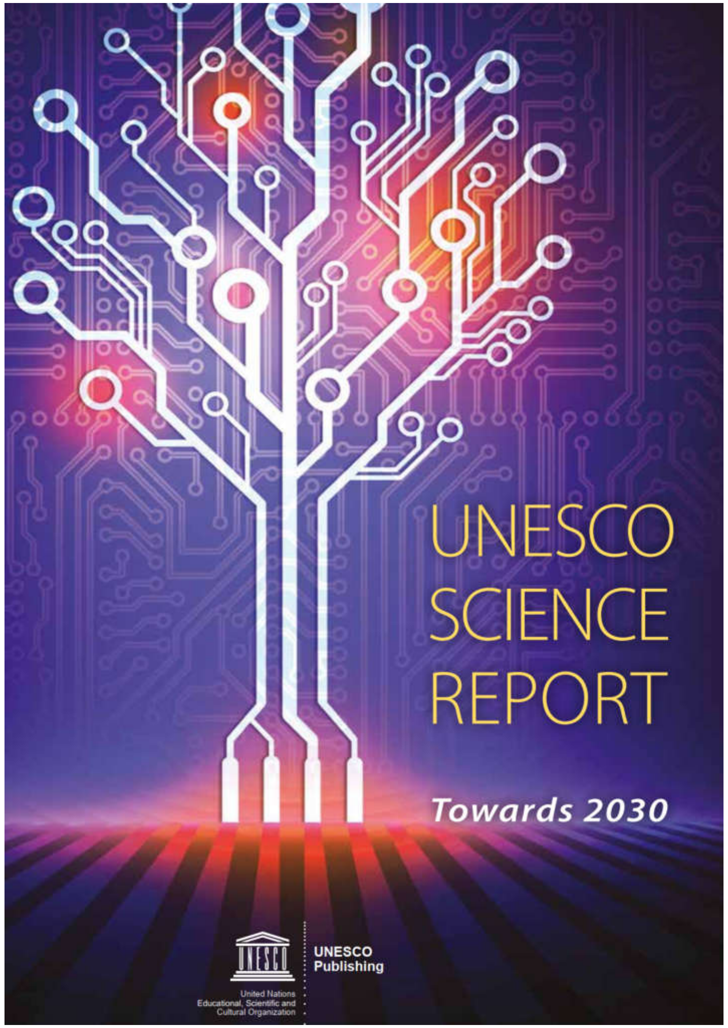
Like 13

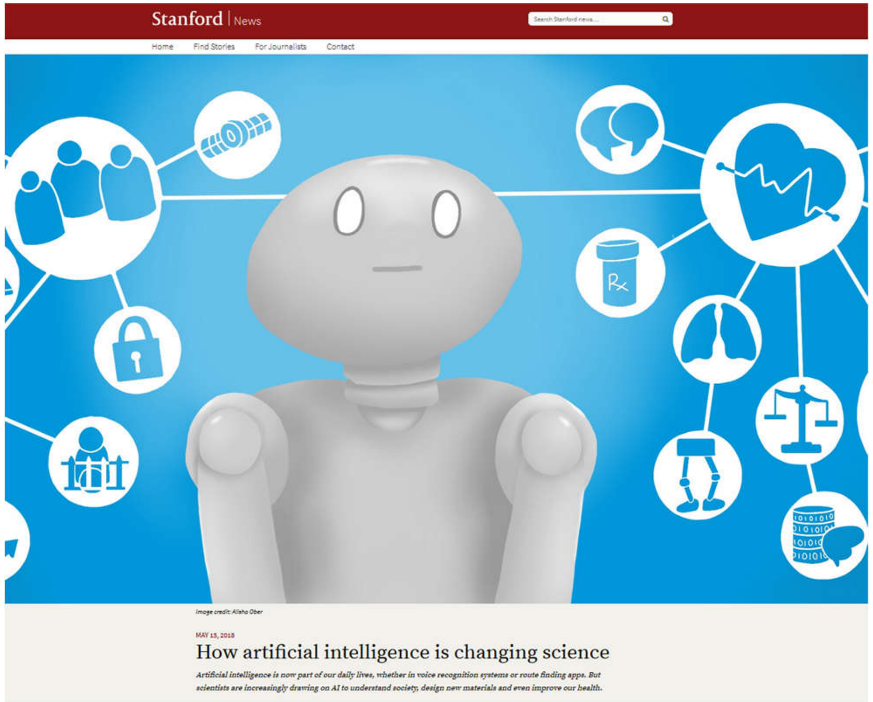
UNESCO SCIENCE REPORT Towards 2030

RELATED LINKS

- Gender and Science
- Disaster Preparedness and Mitigation
- Local and Indigenous Knowledge Systems
- Management of Social Transformations (MOST) Programme
- Science Education
- Engineering
- Science Policy and Capacity Building
- Broadening the Application of the Sustainability Science Approach

https://en.unesco.org/themes/science-society





https://news.stanford.edu/2018/05/15/how-ai-is-changing-science

Friday 14 October lecture on variational inference.

Final Project

In the second half of the course, you will complete a project. The ideal outcome of this project would be a paper that could be submitted to a top-tier machine learning conference such as NIPS, ICML, UAI, AISTATS, or KDD. There are different ways to approach this project, which are discussed in a more comprehensive document that is available from the course website under the Files tab. There are four separate components of the project:

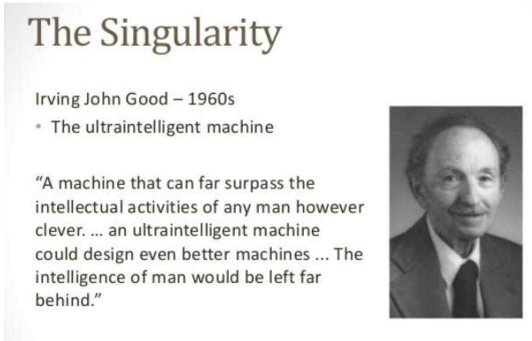
- Progress is driven by the explosion in the availability of **big data** and **low-cost computation.**
- **Health is amongst the biggest challenges**

Jordan, M. I. & Mitchell, T. M. 2015. Machine learning: Trends, perspectives, and prospects. Science, 349, (6245), 255-260.

"An ultra-intelligent machine could design even better machines; there would then unquestionably be an **"intelligence explosion*"** and the intelligence of man would be left far behind …

It is curious that this point is made so seldom … outside of science fiction."

Irving John Good, Trinity College, Oxford, 1965
Colleague of Alan Turing in Bletchley Park



Irving John Good 1966. Speculations Concerning the First Ultraintelligent Machine. Advances in Computers, 6, 31-88, doi: http://dx.doi.org/10.1016/S0065-2458(08)60418-0

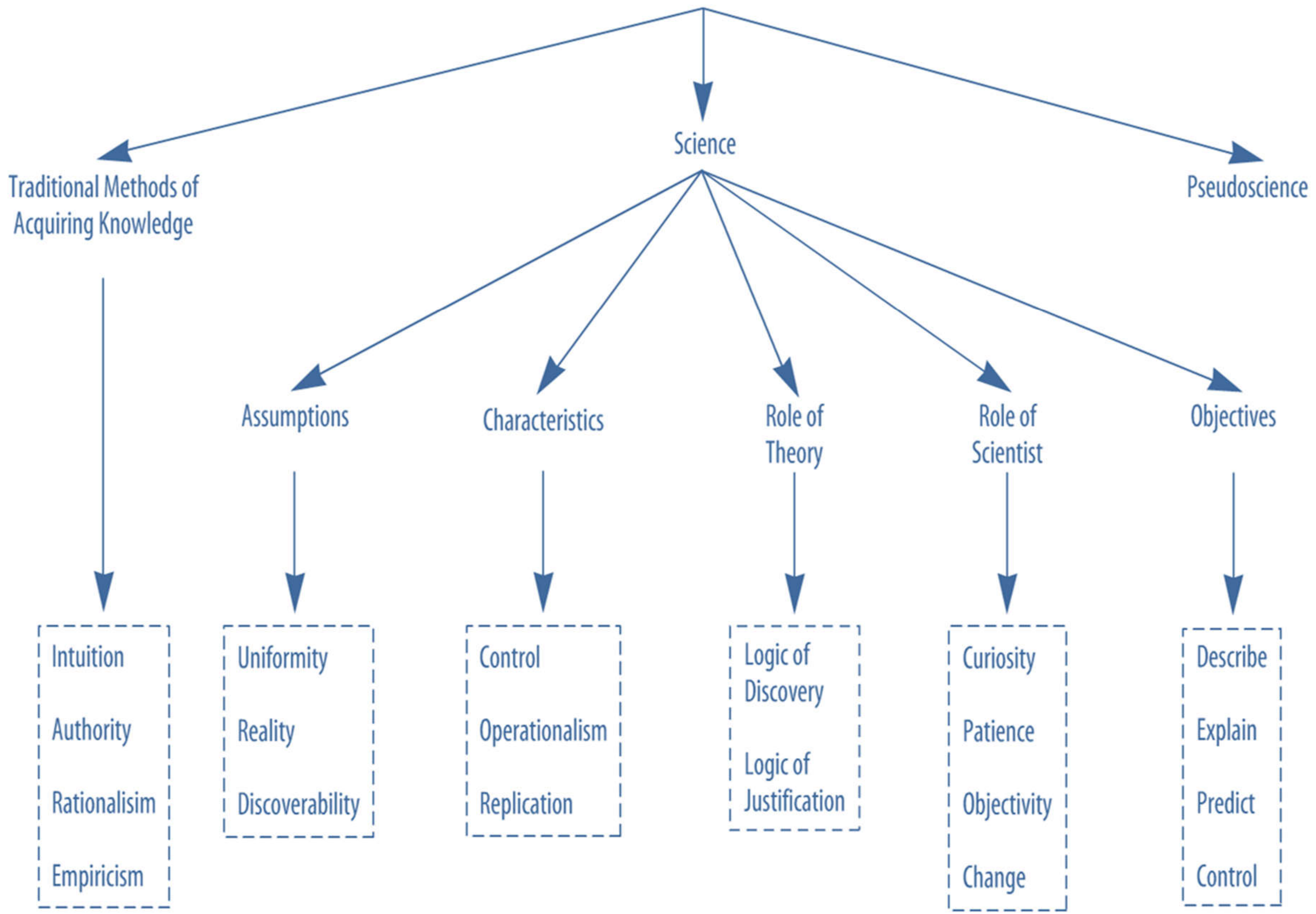
Based on talks given in a Conference on the Conceptual Aspects of Biocommunications, Neuropsychiatric Institute, University of California, Los Angeles, October 1962; and in the Artificial Intelligence Sessions of the Winter General Meetings of the IEEE, January 1963 [1, 46].
The first draft of this monograph was completed in April 1963, and the present slightly amended version in May 1964

*) https://intelligence.org/ie-faq

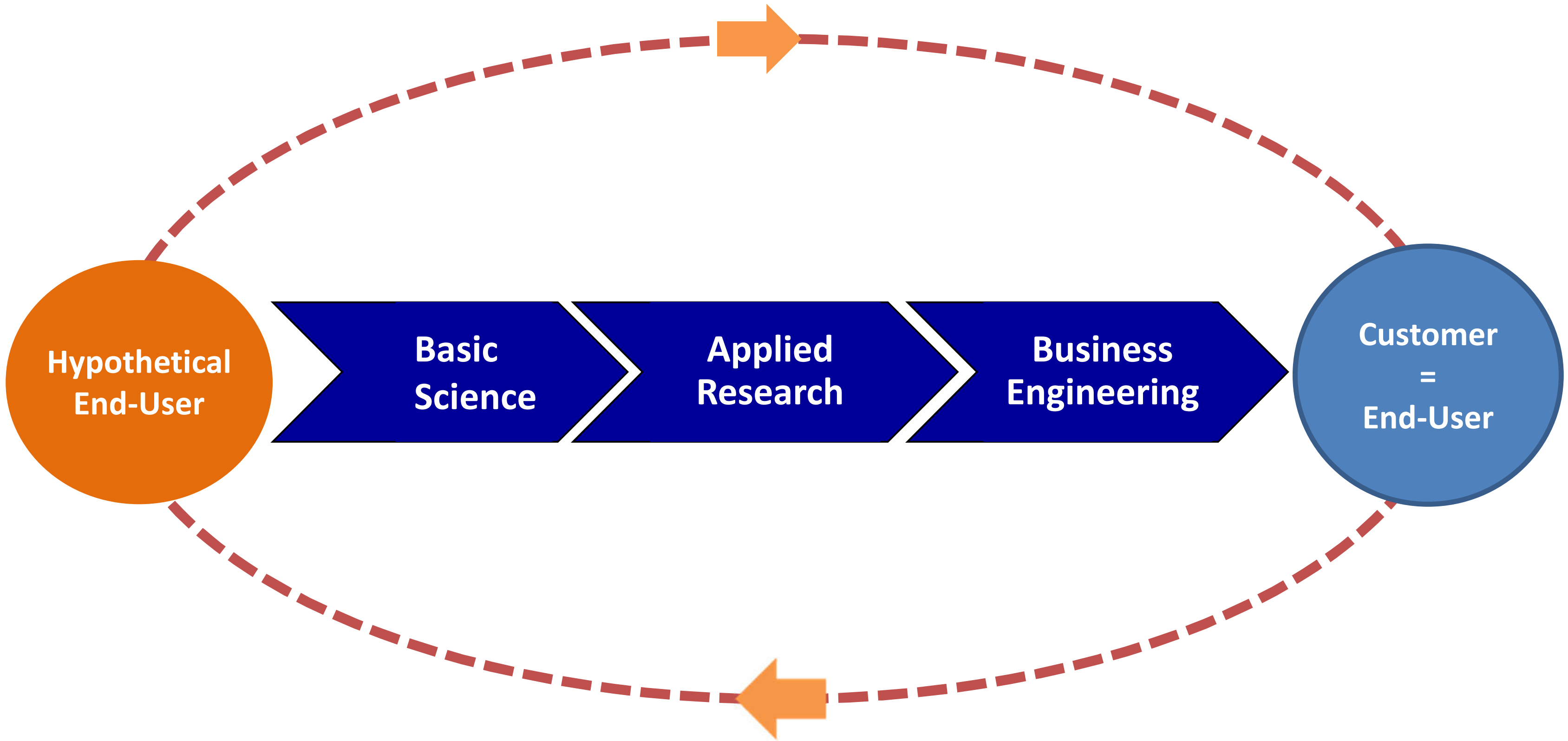
02 What is Science ?
What is Engineering?

- επιστημη (ancient Greek: episteme = Wissen)
- Scientia (latin = Wissen, engl. “knowledge”)
- systematic and objective process to attain and organize new knowledge in the form of testable explanations and predictions about our universe
- Computer Science:
  - A) theoretical
  - B) experimental
  - Three pillars:
    - 1) language = information;
    - 2) process = algorithms;
    - 3) technology = computer (universal programable Machine)

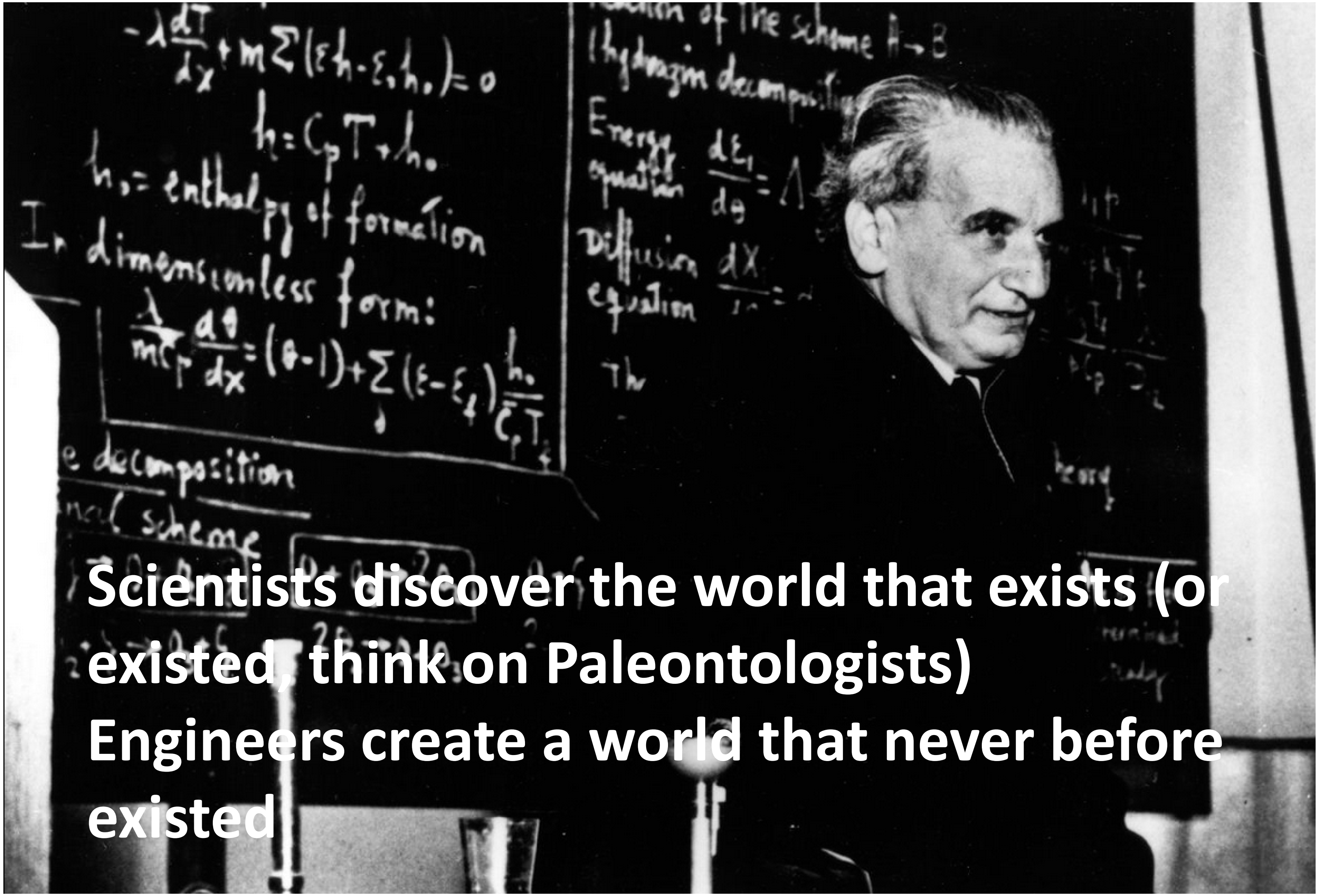
Naïve Assumptions and scientific methods



Science is testing crazy ideas – Engineering is putting these ideas into Business



Holzinger, A. 2011. Successful Management of Research and Development, Norderstedt: BoD.



Scientists discover the world that exists (or existed, think on Paleontologists)
Engineers create a world that never before existed

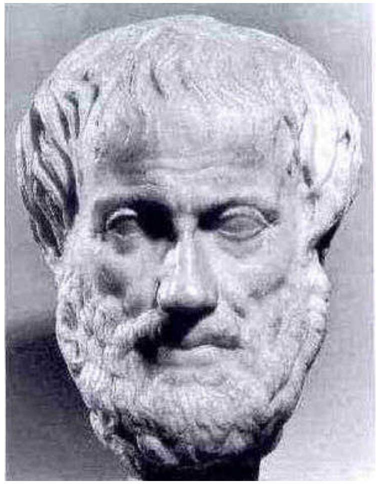
Attributed to Theodore von Karman (1881-1963)

- Computer Science is asking:
  - P = NP ? *)
  - What is information?
  - What is intelligence?
  - What is computable?
- Computer Engineering is asking:
  - (How) can we build (intelligent) information systems (simply)?

  - *) a proof would have profound impact (think on cryptography, TSP, subgraph isomorphism, …

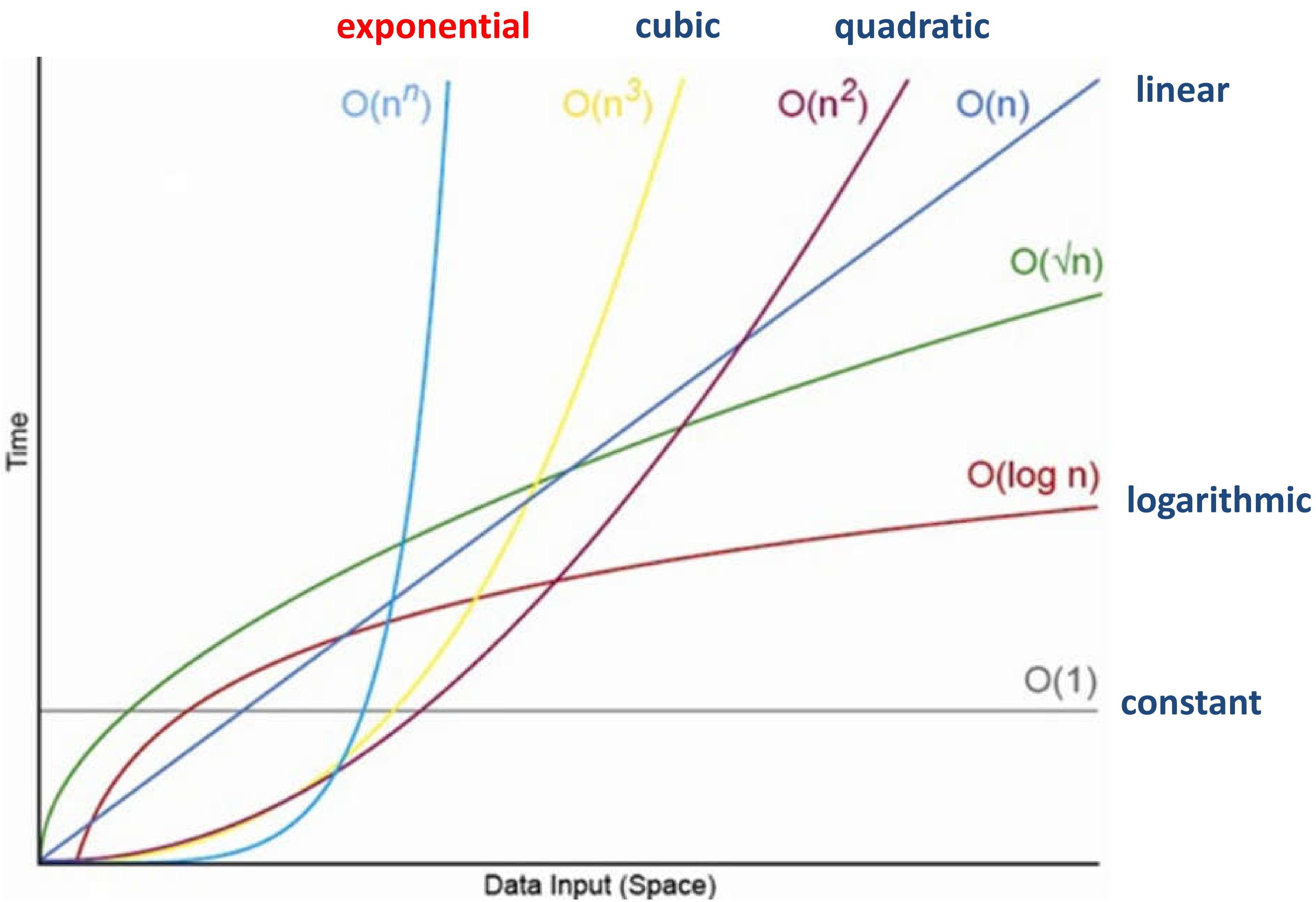
Janet M. Wing 2006. Computational thinking. Communications of the ACM, 49, (3), 33-35, doi:10.1145/1118178.1118215.

According to Aristoteles (384-322)



	Episteme	Techne
Objects	“unchangeable”	Changeable (plastic)
Goal	General knowledge	Specific knowledge
Activities	Building theoria	Building poiesis
Method	Abstraction	Concrete (Modeling)
Process	Conceptualizing	Optimizing
Innovation in form of	Discovery	Invention
Results	Law-like	Rule-like

Complexity Problem: Time versus Space



P versus NP and the Computational Complexity Zoo, please have a look at
https://www.youtube.com/watch?v=YX40hbAHx3s

Who knows this person?

$$\left(-\frac{\hbar^2}{2m}\Delta + U(\vec{r},t)\right)\psi(\vec{r},t) = i\hbar\frac{\partial}{\partial t}\psi(\vec{r},t)$$

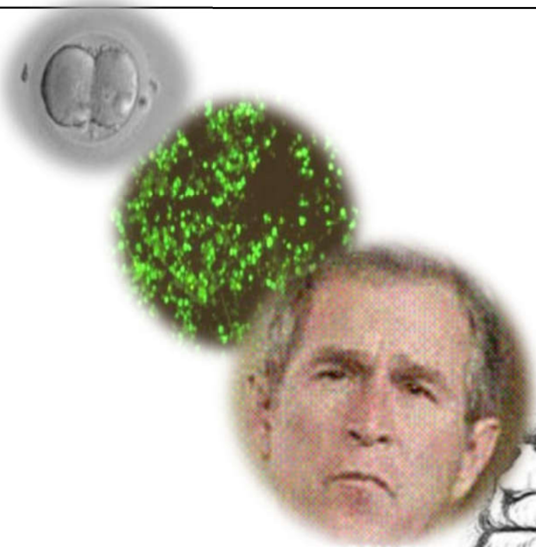
to reproduce …

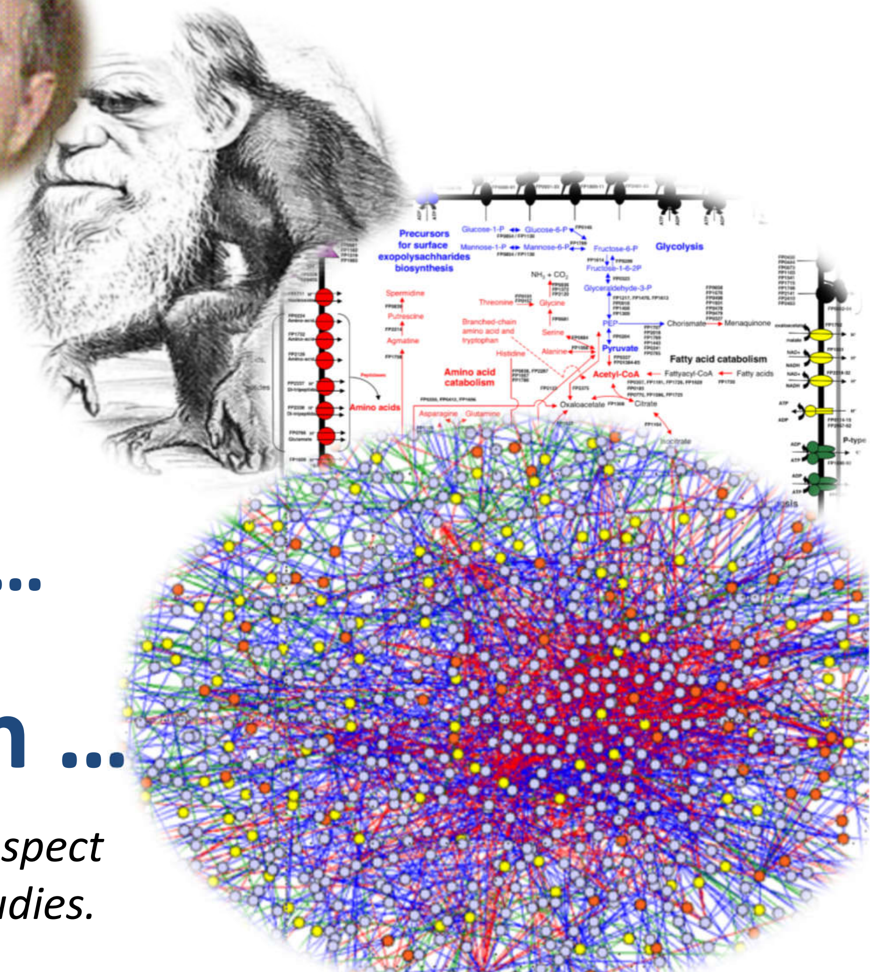
to grow …

to evolve …

to self-replicate …

to generate/utilize energy …

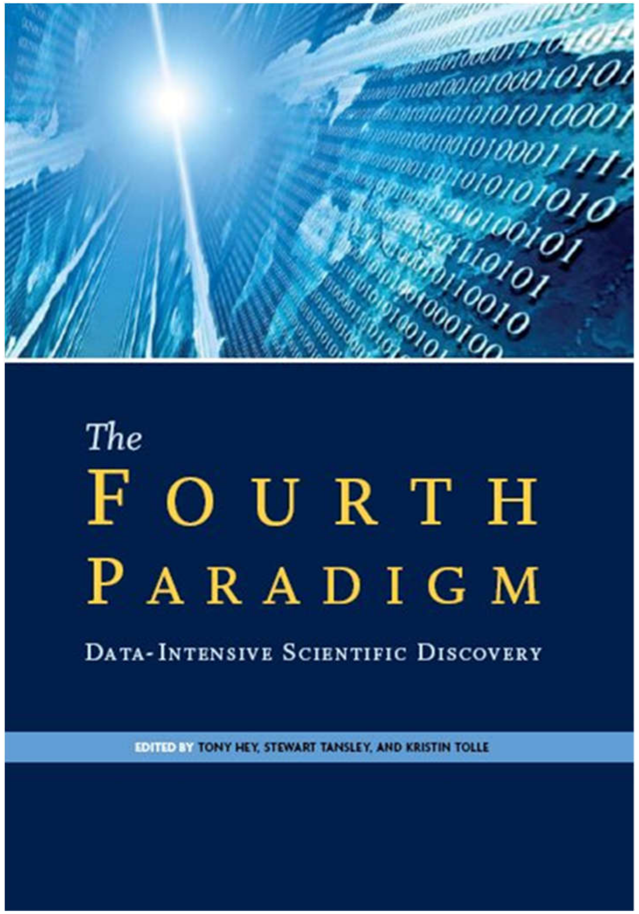




to process information …

Schrödinger, E. (1944) *What Is Life? The Physical Aspect of the Living Cell. Dublin Institute for Advanced Studies.*

The way to eScience – data intensive sciences



Second printing, version 1.1, October 2009.

ISBN 978-0-9825442-0-4

Printed in the United States of America.

Microsoft Research
http://research.microsoft.com

Tony Hey, Stewart Tansley & Kristin Tolle 2009. The fourth paradigm: data-intensive scientific discovery, Redmond (WA), Microsoft Research.

https://www.microsoft.com/en-us/research/publication/fourth-paradigm-data-intensive-scientific-discovery

The information explosion …

practices for the merchant fleet in 1865, *an additional 70 years later.* The total time from Lancaster's definitive demonstration of how to prevent scurvy to adoption across the British Empire was 264 years [2].

The translation of medical discovery to practice has thankfully improved substantially. But a 2003 report from the Institute of Medicine found that the lag between significant discovery and adoption into routine patient care still averages 17 years [3, 4]. This delayed translation of knowledge to clinical care has negative effects on both the cost and the quality of patient care. A nationwide review of 439 quality indicators found that only half of adults receive the care recommended by U.S. national standards [5].

THE IMPACT OF THE INFORMATION EXPLOSION IN MEDICINE

Despite the adoption rate of medical knowledge significantly improving, we face a new challenge due to the exponential increase in the rate of medical knowledge discovery. More than 18 million articles are currently catalogued in the biomedical literature, including over 800,000 added in 2008. The accession rate has doubled every 20 years, and the number of articles per year is expected to surpass 1 million in 2012, as shown in Figure 1.

Translating all of this emerging medical knowledge into practice is a staggering challenge. Five hundred years ago, Leonardo da Vinci could be a painter, engineer, musician, and scientist. One hundred years ago, it is said that a physician might have reasonably expected to know everything in the field of medicine.[2] Today, a typical primary care doctor must stay abreast of approximately 10,000 diseases and syndromes, 3,000 medications, and 1,100 laboratory tests [6]. Research librarians estimate that a physician in just one specialty, epidemiology, needs 21 hours of study per day just to stay current [7]. Faced with this flood of medical information, clinicians routinely fall behind, despite specialization and sub-specialization [8].

The sense of information overload in medicine has been present for surprisingly many years. An 1865 speech by Dr. Henry Noyes to the American Ophthalmologic Society is revealing. He said that "medical men strive manfully to keep up their knowledge of how the world of medicine moves on; but too often they are the first to accuse themselves of being unable to meet the duties of their daily calling…." He went on to say, "The preparatory work in the study of medicine is so great, if adequately done, that but few can spare time for its thorough performance…." [9]

[2] www.medinfo.cam.ac.uk/miu/papers/Hanka/THIM/default.htm

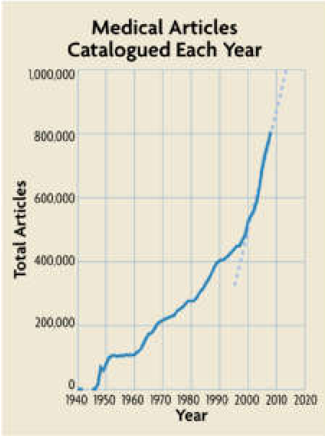


FIGURE 1.
The number of biomedical articles catalogued each year is increasing precipitously and is expected to surpass 1 million in 2012.

COULD KNOWLEDGE ADOPTION IN HEALTHCARE BECOME NEARLY INSTANTANEOUS?

The speed at which definitive medical discoveries have broadly reached medical practice over the last two millennia has progressively increased, as shown in Figure 2 on the next page.

Focusing on the last 150 years, in which the effects of industrialization and the information explosion have been most acute, the trajectory flattens slightly but remains largely linear, as the figure shows. (An asymptotic fit yields an r^2 of 0.73, whereas the linear fit is 0.83.)

Given that even the speed of light is finite, this trend will inevitably be asymptotic to the horizontal axis. Yet, if the linearity can be sufficiently maintained for a while, the next 20 years could emerge as a special time for healthcare *as the translation from medical knowledge discovery to widespread medical practice becomes nearly instantaneous.*

The proximity of this trajectory to the axis occurs around the year 2025. In response to the dramatic computational progress observed with Moore's Law and the growth in parallel and distributed computing architectures, Ray Kurzweil, in *The Singularity Is Near,* predicts that 2045 will be the year of the Singularity, when computers meet or exceed human computational ability and when their ability to recursively improve themselves can lead to an "intelligence explosion" that ultimately affects all aspects of human culture and technology [10]. Mathematics defines a "singularity" as a point at which an object changes its nature so as to attain properties that are no longer the expected norms for that class of object. Today, the dissemination path for medical information is complex and multi-faceted, involving commercials, lectures, brochures, colleagues, and journals. In a world with nearly instantaneous knowledge translation, dissemination paths would become almost entirely digital and direct.

Time from Discovery to Dissemination

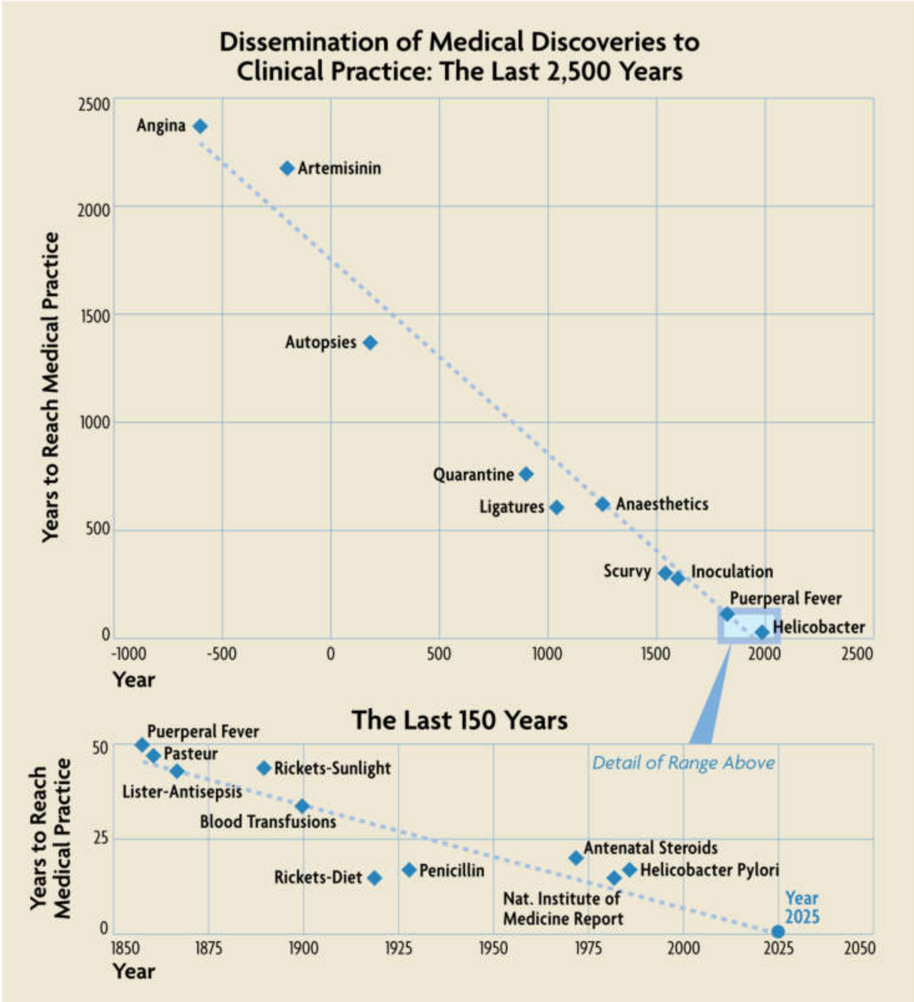


FIGURE 2.

While it took 2,300 years after the first report of angina for the condition to be commonly taught in medical curricula, modern discoveries are being disseminated at an increasingly rapid pace. Focusing on the last 150 years, the trend still appears to be linear, approaching the axis around 2025.

- 1000 years ago – Experimental Science: observing and describing natural phenomena
- 300 years ago – Theoretical Science: Newton’s Laws, Maxwell’s Equation, Einstein, …
- 70 years ago – Computational Science: using universal programmable machines for simulating complex phenomena via math models
- Today – Data-Intensive Science: data captured by instruments, generated by simulations, sensor nets, …

You always get a result – but is it relevant?

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A

MATHEMATICAL, PHYSICAL AND ENGINEERING SCIENCES

Open Access

Check for updates

View PDF

Tools Share

Cite this article

Section

Opinion piece

Big data: the end of the scientific method?

Sauro Succi and Peter V. Coveney

Published: 18 February 2019 https://doi.org/10.1098/rsta.2018.0145

Abstract

"*For it is not the abundance of knowledge, but the interior feeling and taste of things, which is accustomed to satisfy the desire of the soul.* (Saint Ignatius of Loyola)."

We argue that the boldest claims of big data (BD) are in need of revision and toning-down, in view of a few basic lessons learned from the science of complex systems. We point out that, once the most extravagant claims of BD are properly discarded, a synergistic merging of BD with big theory offers considerable potential to spawn a new scientific paradigm capable of overcoming some of the major barriers confronted by the modern scientific method originating with Galileo. These obstacles are due to the presence of nonlinearity, non-locality and hyperdimensions which one encounters frequently in multi-scale modelling of complex systems.

This article is part of the theme issue 'Multiscale modelling, simulation and computing: from the desktop to the exascale'.

Sauro Succi & Peter V. Coveney 2019. Big data: the end of the scientific method? Philosophical Transactions of the Royal Society A, 377, (2142), 20180145, doi:10.1098/rsta.2018.0145.

Available online: https://royalsocietypublishing.org/doi/full/10.1098/rsta.2018.0145

03 Basics of the Theory of Science (Wissenschaftstheorie)

- **Classic Newtonian** approach:
  - Ask question > develop theory > form a hypothesis to proof/disproof theory > conduct experiments > compare data with hypothesis > accept/reject theory
- **Computer Science** approach:
  - Find open problems to solve > form hypothesis how to solve the problem > experiment > evaluate > present new solution to the problem
- **Modern Machine Learning** approach:
  - Setting up experiments to answer questions including: How does model *m* perform on data *d* from domain *D*? Which of these models have the best performance? Much is feature engineering and precision and recall are your best friends! Now questions of causality to answer the questions of why are becoming important!

General overview of the classic approaches

	Objects	Primary Method
Logic & Mathematics	Simple abstract objects: numbers, propositions, …	Deduction
Natural Sciences	Natural objects: physical objects, fields, organisms, …	Hypothetico-deductive
Social Sciences	Social objects: human individuals, collectives, society, …	Hypothetico-deductive & Hermeneutics
Humanities (e.g. Law)	Complex cultural objects: human ideas, principles, actions, relationships, language, artefacts, …	Hermeneutics

Remember: Reasoning = “Sensemaking”

- **Deductive Reasoning** = Hypothesis > Observations > Logical Conclusions
  - DANGER: Hypothesis must be correct! DR defines whether the truth of a conclusion can be determined for that rule, based on the truth of premises: A=B, B=C, therefore A=C
- **Inductive reasoning** = makes broad generalizations from specific observations
  - DANGER: allows a conclusion to be false if the premises are true
  - generate hypotheses and use DR for answering specific questions
- **Abductive reasoning** = inference = to get the best explanation from an incomplete set of preconditions.
  - Given a true conclusion and a rule, it attempts to select some possible premises that, if true also, may support the conclusion, though not uniquely.
  - Example: "When it rains, the grass gets wet. The grass is wet. Therefore, it might have rained." This kind of reasoning can be used to develop a hypothesis, which in turn can be tested by additional reasoning or data.

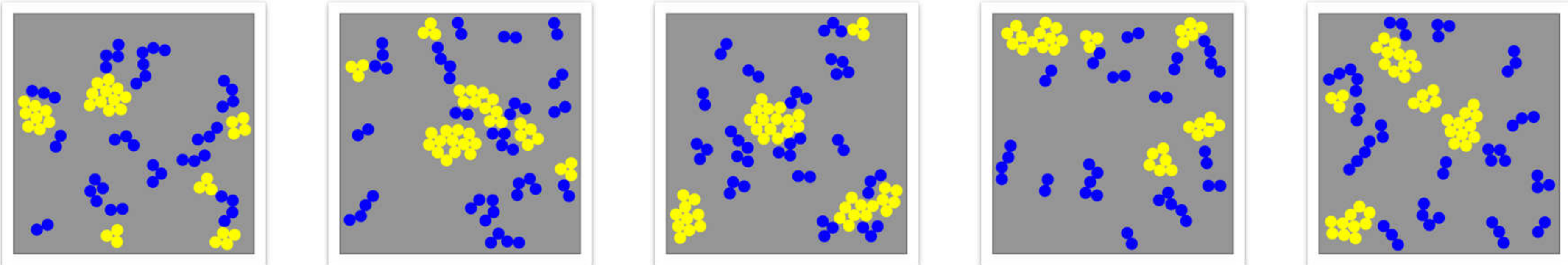
Important Definition: Ground truth

- := information provided by direct observation (empirical evidence) in contrast to information provided by inference
  - Empirical evidence = information acquired by observation or by experimentation in order to verify the truth (fit to reality) or falsify (non-fit to reality).
  - Empirical inference = drawing conclusions from empirical data (observations, measurements)
  - Causal inference = drawing a conclusion about a causal connection based on the conditions of the occurrence of an effect.
    - Causal inference is an example of causal reasoning.

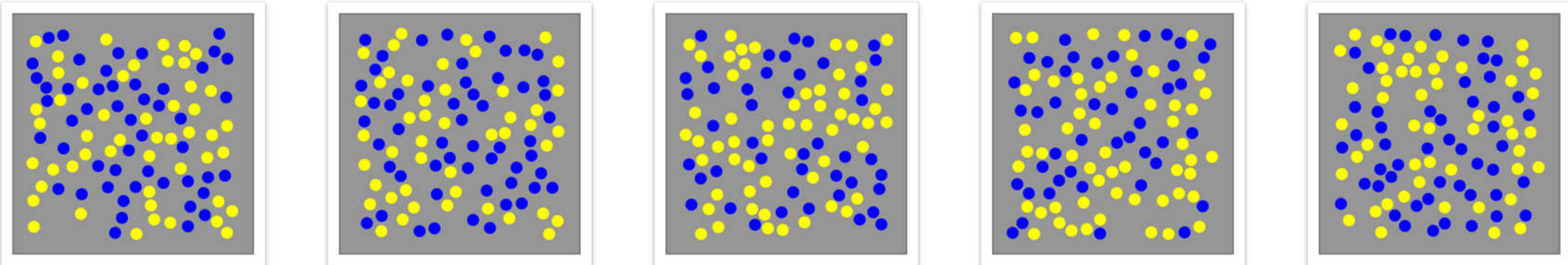
Thus, the KandinskyPatterns are so important!

https://human-centered.ai/project/kandinsky-patterns

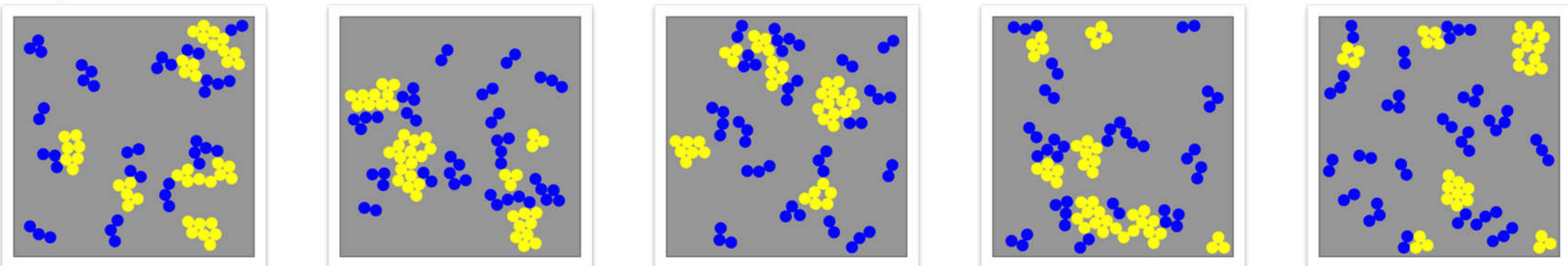
A) True



B) False



C) Counterfactual



04 Hypothetico-Deductive Method

The scientific method is strictly objective

https://www.youtube.com/watch?v=wf-sGqBsWv4

- The scientific method is the logical scheme used by scientists searching for answers to the questions posed within science.
- The Scientific method is used to produce scientific theories, including both scientific meta-theories (theories about theories) as well as the theories used to design the tools for producing theories (instruments, algorithms, etc).

- Observe an event
- Ask a question ("Originäre Fragestellung") and check with the state-of-the-art whether and to what extent this question has already been answered!!
- Develop a model (or hypothesis) which makes a prediction to explain this event
- Test your prediction with (new) data
- Observe the result
- Proof the hypothesis or revise appropriately
- repeat as long as needed
- A successful model (or hypothesis) becomes a scientific theory !

What is the (expected) contribution for the scientific community ?

and playing the game Go.[26] Yet machine learning algorithms that exceed human performance in naturally occurring scenarios are often seen as failing dramatically when an adversary is able to modify their input data even subtly. Machine learning is already used for many highly important applications and will be used in even more of even greater importance in the near future. Search algorithms, automated financial trading algorithms, data analytics, autonomous vehicles, and malware detection are all critically dependent on the underlying machine learning algorithms that interpret their respective domain inputs to provide intelligent outputs that facilitate the decision-making process of users or automated

56 COMMUNICATIONS OF THE ACM | JULY 2018 | VOL. 61 | NO. 7

key insights

- Machine learning has traditionally been developed following the assumption that the environment is benign during both training and evaluation of the model; while useful for designing effective algorithms, this implicitly rules out the possibility that an adversary could alter the distribution at either training time or test time.
- In the context of adversarial inputs at test time, few strong countermeasures exist for the many attacks that have been demonstrated.
- To end the arms race between attackers and defenders, we suggest building more tools for verifying machine learning models; unlike current testing practices, this could help defenders eventually gain a fundamental advantage.

Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images

Duc Fehr, Harini Veeraraghavan, Andreas Wibmer, Tatsuo Gondo, Kazuhiro Matsumoto, Herbert Alberto Vargas, Evis Sala, Hedvig Hricak, and Joseph O. Deasy

ELSEVIER
Information Sciences
Volume 295, 20 February 2015, Pages 395-406

A rapid learning algorithm for vehicle classification

Xuezhi Wen, Ling Shao, Yu Xue, Wei Fang

Show more

https://doi.org/10.1016/j.ins.2014.10.040

Get rights and content

Highlights

- A fast learning algorithm is introduced for real-time vehicle classification.
- A fast feature selection method for AdaBoost is presented by combining a sample's feature value with its class label.
- A rapid incremental learning algorithm of AdaBoost is designed.

ELSEVIER
Medical Image Analysis
Volume 35, January 2017, Pages 18-31

Brain tumor segmentation with Deep Neural Networks

Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, Hugo Larochelle

Show more

https://doi.org/10.1016/j.media.2016.05.004

Get rights and content

Highlights

- A fast and accurate fully automatic method for brain tumor segmentation which is competitive both in terms of accuracy and speed compared to the state of the art.
- The method is based on deep neural networks (DNN) and learns features that are specific to brain tumor segmentation.
- We present a new DNN architecture which exploits both local features as well as more global contextual features simultaneously.
- Using a GPU implementation and a convolutional output layer, the model is an order of magnitude faster than other state of the art methods.

Social big data: Recent achievements and new challenges

Gema Bello-Orgaz, Jason J. Jung, David Camacho

Show more

https://doi.org/10.1016/j.inffus.2015.08.005

Get rights and content

Highlights

- The paper presents the methodologies on information fusion for social media.
- The methodologies, frameworks, and software used to work with big data are given.
- The state of the art in the data analytic techniques on social big data is provided.
- Social big data applications for various domains are described and analyzed.

PLOS ONE

PUBLISH ABOUT BROWSE SEARCH

OPEN ACCESS PEER-REVIEWED
RESEARCH ARTICLE

On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation

Sebastian Bach, Alexander Binder, Grégoire Montavon, Frederick Klauschen, Klaus-Robert Müller, Wojciech Samek

Published: July 10, 2015 • https://doi.org/10.1371/journal.pone.0130140

Article | Authors | Metrics | Comments | Media Coverage

Download PDF | Print | Share

Abstract

Understanding and interpreting classification decisions of automated image classification systems is of high value in many applications, as it allows to verify the reasoning of the system and provides additional information to the human expert. Although machine learning methods are solving very successfully a plethora of tasks, they have in most cases the disadvantage of acting as a black box, not providing any information about what made them arrive at a particular decision. This work proposes a general solution to the problem of understanding classification decisions by pixel-wise decomposition of nonlinear classifiers. We introduce a methodology that allows to visualize the contributions of single pixels to predictions for kernel-based classifiers over Bag of Words features and for multilayered neural networks. These pixel contributions can be visualized as heatmaps and are provided to a human expert who can intuitively not only verify the validity of the classification decision, but also focus further analysis on regions of potential interest. We evaluate our method for classifiers trained on PASCAL VOC 2009 images, synthetic image data containing geometric shapes, the MNIST handwritten digits data set and for the pre-trained ImageNet model available as part of the Caffe open source package.

Figures

Subject Areas: Neurons, Neural networks, Kernel functions, Algorithms, Coding mechanisms, Imaging techniques, Kernel methods, Vision

THE ROYAL SOCIETY PUBLISHING | All Journals

Home Content Information for About us Sign up Submit

Section: Abstract; 1. Introduction to deep learning; 2. Deep learning and patient categorization; 3. Deep learning to study the fundamental biological processes underlying human disease; 4. The impact of deep learning in treating disease and developing new treatments; 5. Discussion; 6. Conclusion; 7. Methods

Authors' contributions

We created an open repository on the GitHub version control platform (greenelab/deep-review) [555]. Here, we engaged with numerous authors from papers within and outside of the area. The manuscript was drafted via GitHub commits by 36 individuals who met the ICMJE standards of authorship. These were individuals who contributed to the review of the literature; drafted the manuscript or provided substantial critical revisions; approved the final manuscript draft; and agreed to be accountable in all aspects of the work. Individuals who did not contribute in all of these ways, but who did participate, are acknowledged below. We grouped authors into the following four classes of approximately equal contributions and randomly ordered authors within each contribution class. Drafted multiple sub-sections along with extensive editing, pull request reviews or discussion: A.A.K., B.K.B., B.T.D., D.S.H., E.F., G.P.W., M.M.H., M.Z., P.A. and T.C. Drafted one or more sub-sections: A.E.C., A.M.A., A.S., B.J.L., C.A.L., E.M.C., G.L.R., J.I., J.L., J.X., S.C.T., S.W., W.X. and Z.L. Revised specific sub-sections or supervised drafting one or more sub-sections: A.H., A.K., D.D., D.J.H., L.K.W., M.H.S.S., S.J.S., S.M.B., Y.P. and Y.Q. Drafted sub-sections, edited the manuscript, reviewed pull requests and coordinated co-authors: A.G. and C.S.G.

Competing interests

A.K. is on the Advisory Board of Deep Genomics Inc. E.F. is a full-time employee of GlaxoSmithKline. The remaining authors have no competing interests to declare.

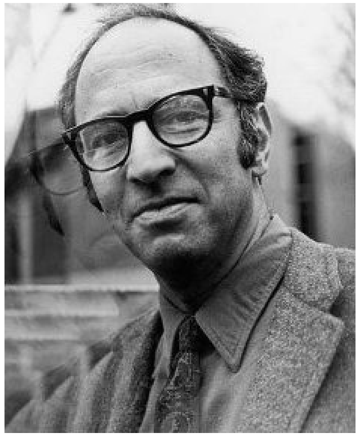
Steps (simplified)

1. Formulate a research question in the context of existing knowledge (theory & observations) – you must know who did what when and where!
2. Formulate a hypothesis as a tentative answer to this question
3. Deduce consequences and make predictions
4. Test the hypothesis in a specific experiment/theory field. The new hypothesis must prove to fit-in the existing world-view, think about what Sir Karl Popper said!

Warning

- In case the hypothesis leads to contradictions and demands a radical change in the existing theoretical background, test it carefully again!
- If you succeed and it replaces the existing scientific paradigm – this is called “scientific revolution” and it happens very rarely and cannot be planned …
- Repeat the process with modifications of the hypothesis until you reach an agreement which leads to a satisfiable result
- If you find major discrepancies, you must start the process from the beginning, or you state an alternative research question!

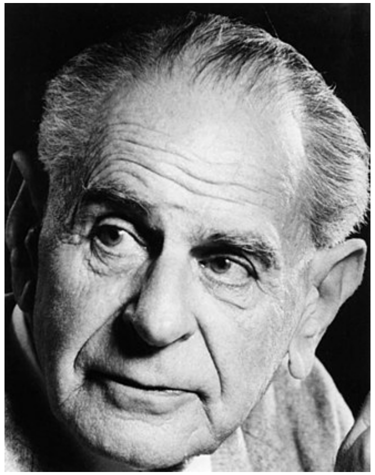
- When consistency is obtained the hypothesis becomes a theory and provides a coherent set of propositions that define a new class of phenomena or a new theoretical concept.
- The results have to be published and is subject of process of ”natural selection” among competing theories … reviewer give you a hard time!
- A theory is then becoming a framework within which observations/theoretical facts are explained and predictions can be made.



- Science undergoes always periodic paradigmatic changes
- These paradigm shifts open up new approaches
- Scientists can (of course) never separate their subjective perspective from their work
- thus, our comprehension of science can never rely on full objectivity according to Kuhn

https://en.wikipedia.org/wiki/Commensurability_(philosophy_of_science)
For further information on the work of Thomas Kuhn et al.

Karl Popper (1902-1994) on positivism



- There is no logical path leading to [the highly universal laws of science]. They can only be reached by intuition, based upon something like an intellectual love of the objects of experience'.
- Science starts with problems (yes, we engineers call it challenges !).
- Coping with a problem, the scientist makes observations.
- Observations are selectively designed to test if a theory functions as a satisfactory solution to a given problem.
- Read more here: http://plato.stanford.edu/entries/popper

Karl Popper (contd) on Falsification as principle

- Science is interested in universal affirmative conclusions.
- However, such conclusions could never be verified.
- But, they could be falsified by the discovery of a counterexample!
- Science should aim not to verify or confirm hypotheses but to falsify them.
- According to Popper, there can be never a confirmation (Bestätigung) of a hypothesis.
- It can only be a corroboration (Bekräftigung, Erhärtung)

Richard Feynman on the scientific method

https://www.youtube.com/watch?v=0KmimDq4cSU

05 Occam's Razor

- Occam's razor (novacula Occami; or law of parsimony: Latin: lex parsimoniae)
- is the problem-solving principle that states "Entities should not be multiplied without necessity.“
- “Pluralitas non est ponenda sine necessitate”
- “All else equal – prefer the simplest theory”
- "The simplest solution is most likely the right one."
- Occam's razor says that when presented with competing hypotheses that make the same predictions, one should select the solution with the fewest assumptions

William of Ockham, or Occam.
This image is in the public domain

Problem is Complexity:
Ad-hoc hypotheses
Multiple mechanisms
Coincidences
Many free parameters, …

Parsimony is important for explainable AI

- Explanation is difficult:
  - A) Is the minimum explanation the simplest ?
  - B) Is the simplest explanation the best explanation ?
  - C) When is it enough? (“where is the saturation point”)
  - D) How can an explanation be adapted to different previous knowledge?

Pedro Domingos 1999. The Role of Occam's Razor in Knowledge Discovery. Data Mining and Knowledge Discovery, 3, (4), 409-425, doi:10.1023/a:1009868929893.

Anselm Blumer, Andrzej Ehrenfeucht, David Haussler & Manfred K Warmuth 1987. Occam's razor. Information processing letters, 24, (6), 377-380.

- Note: it means to prefer the most obvious explanation, e.g. you cheese is vanished:
- A) most likely it has been eaten by a mouse
- B) most unlikely it has been taken by a Marsian
- Danger: “best explanation” in what sense?
- Occam originally emphasized that you should always take an explanation with the fewest assumptions
- Counterexample: Quantum Mechanics

"Nature operates in the shortest way possible" – Aristotle.

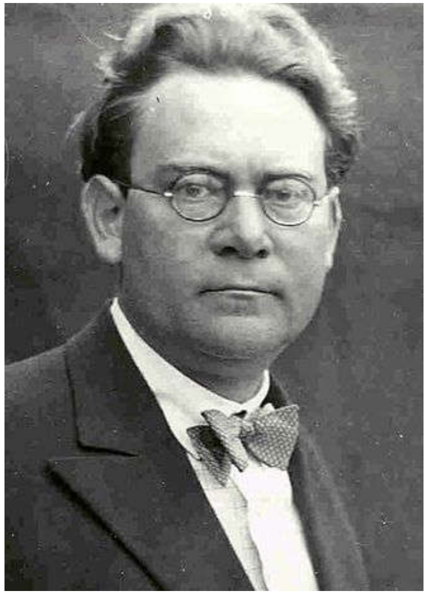
"Entities should not be multiplied without necessity" – William of Occam.

"Everything should be made as simple as possible, but not simpler" – Albert Einstein.

06 Reichenbach's Principle

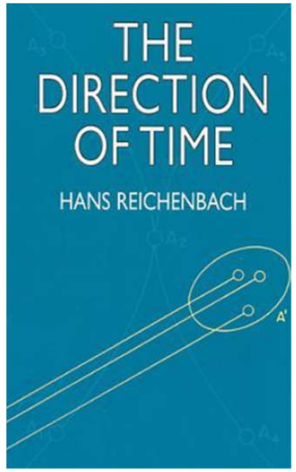
Hans Reichenbach (1891-1953)

Proponent of logical empiricism.



- We can use probability to decide if A is a possible cause of B.
- If P(A&B) > P(A)P(B) we say that A and B are (positively) correlated. Then A could be a cause of B.
- The condition is equivalent to P(B|A)>P(B).
- The condition is symmetric in A and B: If A is a possible cause of B, then B is a possible cause of A.
- "Which came first? The hen or the egg."

Reichenbach’s common cause principle



”If an improbable coincidence has occurred, there must exist a common cause” (p. 157)

Two examples: ”Suppose both lamps in a room go out suddenly. We regard it as improbable that by chance both bulbs burned out at the same time and look for a burned out fuse or some other interruption of the common power supply. The improbable coincidence is thus explained as the product of a common cause.”

”Or suppose several actors in a stage play fall ill showing symptoms of food poisoning. We assume that the poisoned food stems from the same source – for instance, that it was contained in a common meal – and then look for an explanation of the coincidence in terms of a common cause.”

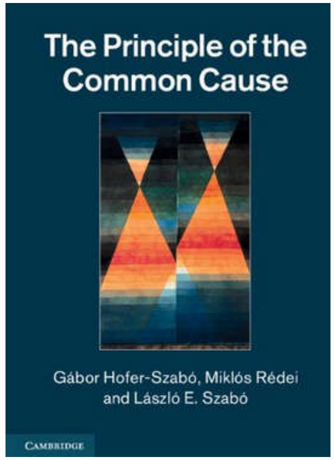
Classical probability measure space: (Ω, Σ, p)

Positive correlation: A, $B \in \Sigma$

$$p(AB) > p(A)\,p(B)$$

Reichenbachian common cause: $C \in \Sigma$

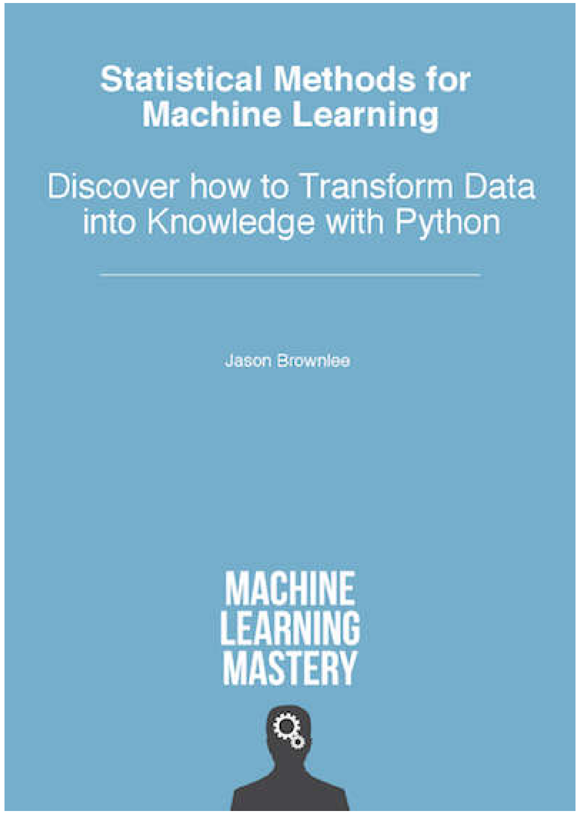
$$\begin{aligned} p(AB|C) &= p(A|C)p(B|C) \\ p(AB|C^{\perp}) &= p(A|C^{\perp})p(B|C^{\perp}) \\ p(A|C) &> p(A|C^{\perp}) \\ p(B|C) &> p(B|C^{\perp}) \end{aligned}$$

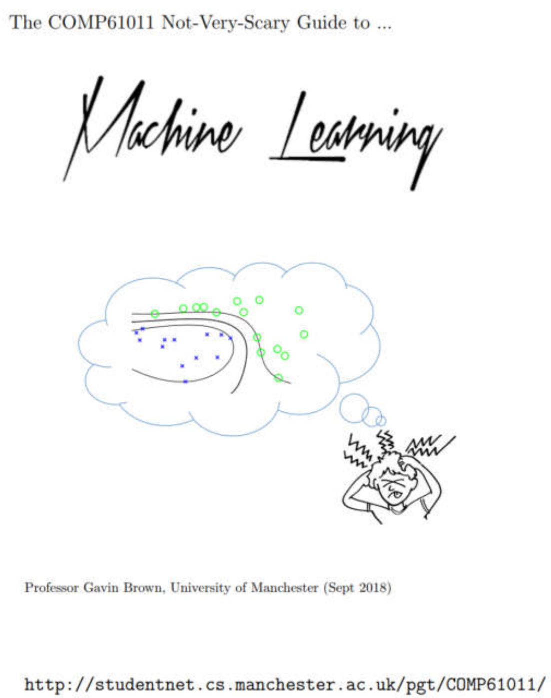


- **Reichenbach's common cause principle:** Assume that *X not independent Y* .
- Then
  X causes *Y* ,
  Y causes *X*,
  there is a hidden common cause or combination.
  For example:
- X = You like this lecture hall
  Y = You like this theory
  Z = You attend this course

07 Experiments in Machine Learning

Some excellent resources …





https://machinelearningmastery.com

http://syllabus.cs.manchester.ac.uk/pgt/2019/COMP61011

Why is measuring important?

"If you can't measure it,

nor assign it an exact numerical value, nor express it in numbers,

then your knowledge is of a meager and unsatisfactory kind"

(attributed to William Thomson (1824-1907), aka Lord Kelvin)

What to measure?
How to measure?
How to interpret?

Systematic Experimentation

- The answers to the question of which algorithm works best on your specific data set or problem, or which input features to use, can only be found through rigorous experiments.
- This is because many ML algorithms are too complex for formal analysis, at least at the level of generality assumed by most theoretical treatments. As a result, empirical studies of the behaviour of machine learning algorithms must retain a central role.

Excellent source: https://machinelearningmastery.com/controlled-experiments-in-machine-learning

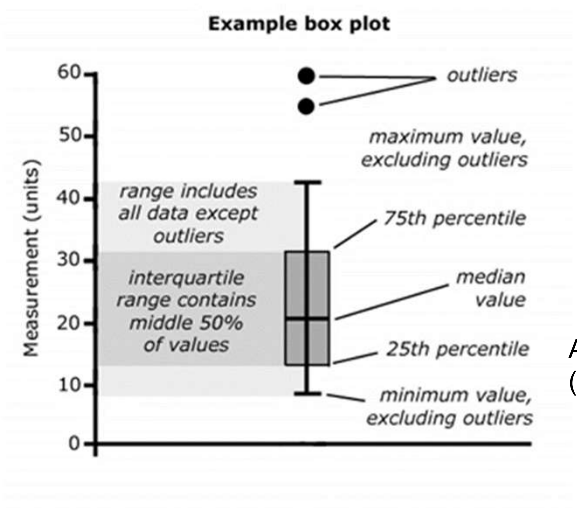
Experimentation is key in applied machine learning

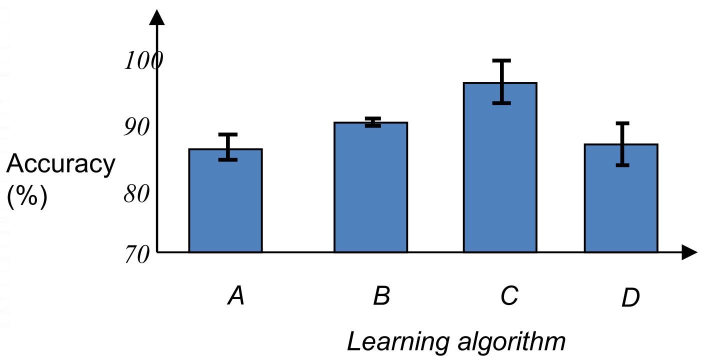
- This is a challenge for beginners who must learn some rigor.
- Three types of experiments:
- Choose-Features Experiments. When determining what data features (input variables) are most relevant to a model, the independent variables may be the input features and the dependent variable might be the estimated skill of the model on unseen data.
- Tune-Model Experiments. When tuning a machine learning model, the independent variables may be the hyperparameters of the learning algorithm and the dependent variable might be the estimated skill of the model on unseen data.
- Compare-Models Experiments. When comparing the performance of machine learning models, the independent variables may be the learning algorithms themselves with a specific configuration and the dependent variable is the estimated skill of the model on unseen data.
- What makes the experimental focus of applied machine learning so exciting is two-fold:
  - 1) Discovery. You can discover what works best for your specific problem and data.
  - 2) Contribution. You can make broader discoveries in the field, without any specialized knowledge other than rigorous and systematic experimentation.

Excellent source: https://machinelearningmastery.com/controlled-experiments-in-machine-learning

NOTE: Learning Algorithms depend on your test data

- Cross-Validation is a way to monitor stability
- Check always confidence intervals
- ROC-Analysis, particularly for imbalanced data
  - (We rarely have I.I.D. data !!!)
  - Check: Accuracy, Training Time, Space complexity (how much memory is needed), Interpretability = how can we explain why it does what it does!
  - Mutual Information, see MacKay, Section 44.5

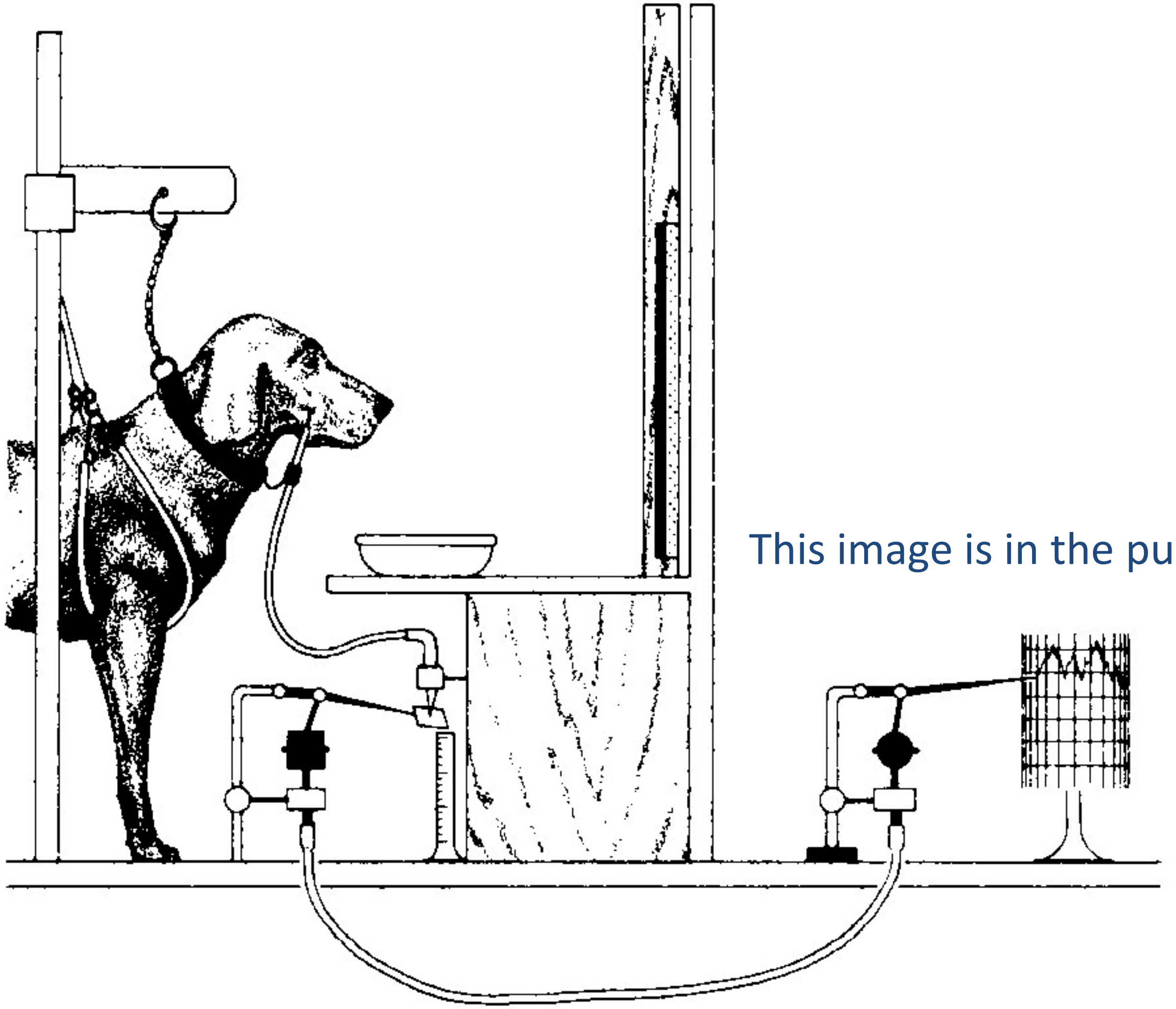




Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

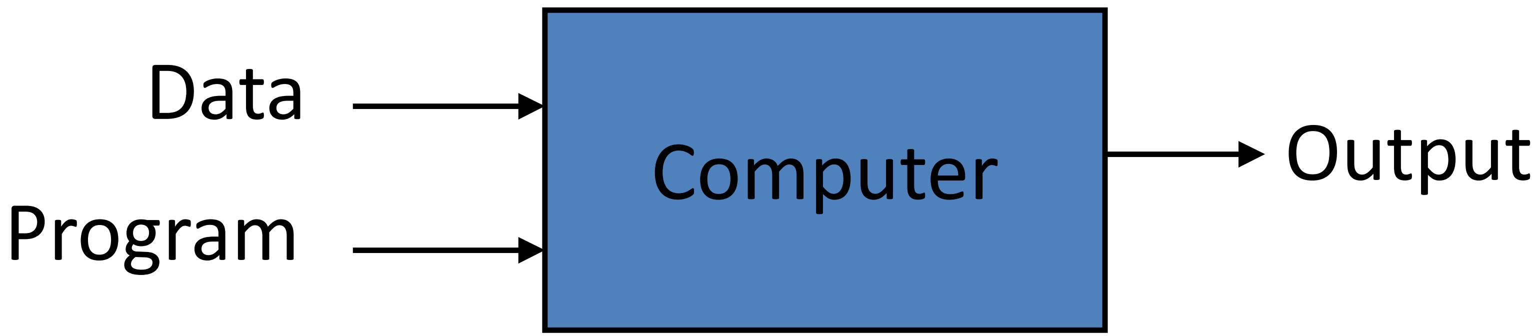
Classical Experiment



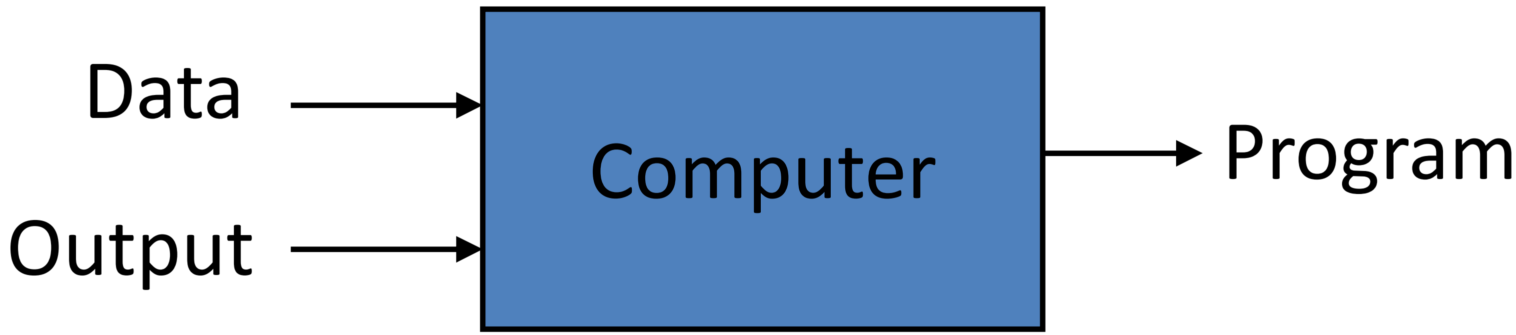
This image is in the public domain

- Machine learning experiments pose questions about models that we try to answer by means of measurements on data. We may ask questions including:
- Which learning algorithm provides the best model for data from Domain D?
- How does the model m perform on data from D?
- What model has the best performance on D?
- How do you benchmark? What is the ground truth?

Traditional Programming



Machine Learning = Learning from Data



There is no free lunch!

Wolpert, D. H. & Macready, W. G. 1997. No free lunch theorems for optimization. IEEE Transactions on Evolutionary Computation, 1, (1), 67-82.

- Scalability
- Predictive accuracy = Hit rate
- Weighted (cost-sensitive) accuracy
- Speed (on model building and predicting)
- Robustness (one weakness in iML-approach)
- Precision/Recall (F-Measure, Break Even Point)
- Area under the ROC (see next slides)

Japkowicz, N. & Shah, M. 2011. Evaluating learning algorithms: a classification perspective, Cambridge University Press.

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems

FYI: Datasets for benchmarking purposes

- There are many datasets for testing machine learning algorithms, just some examples:
- https://www.kaggle.com
- http://archive.ics.uci.edu/ml/datasets.html (UCI Machine Learning Repository)
- http://image-net.org
- http://yann.lecun.com/exdb/mnist (handwritten digit database)
- https://data.medicare.gov/

http://hci-kdd.org/open-data-sets/

- Question: is 99% accuracy good?
- Answer: It depends on the problem!

68

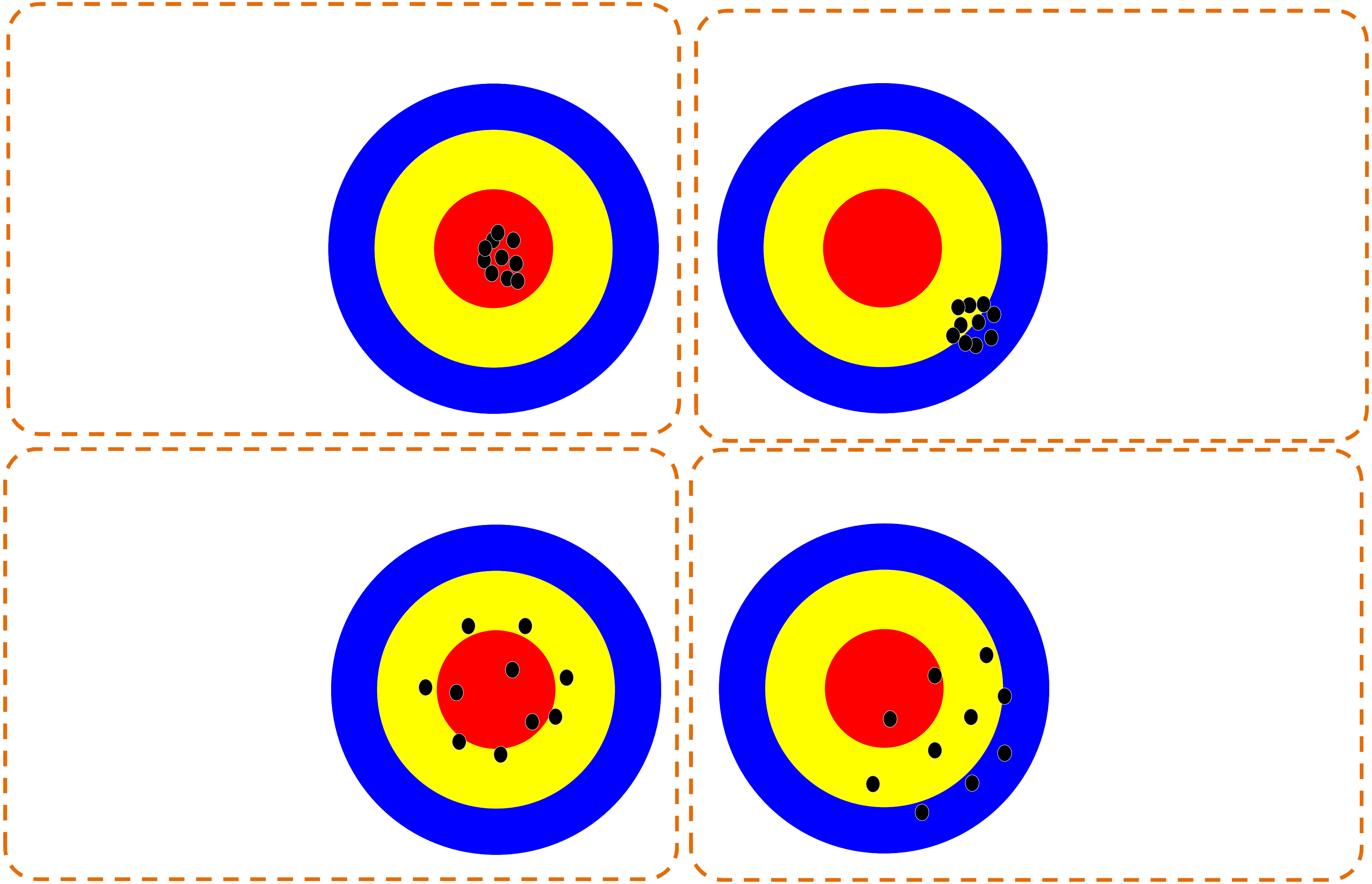
Please always remember these four terms:

- **Accuracy** = error rate of correct/incorrect predictions made by the model over a data set (cf. coverage).
- **Precision** = precision (positive predictive value) is the fraction of retrieved instances that are relevant, while **Recall** (aka sensitivity) is the fraction of relevant instances that are retrieved
- **Reliability** = basically the "consistency" or "repeatability"
- **Validity** = generally, to get valid conclusions

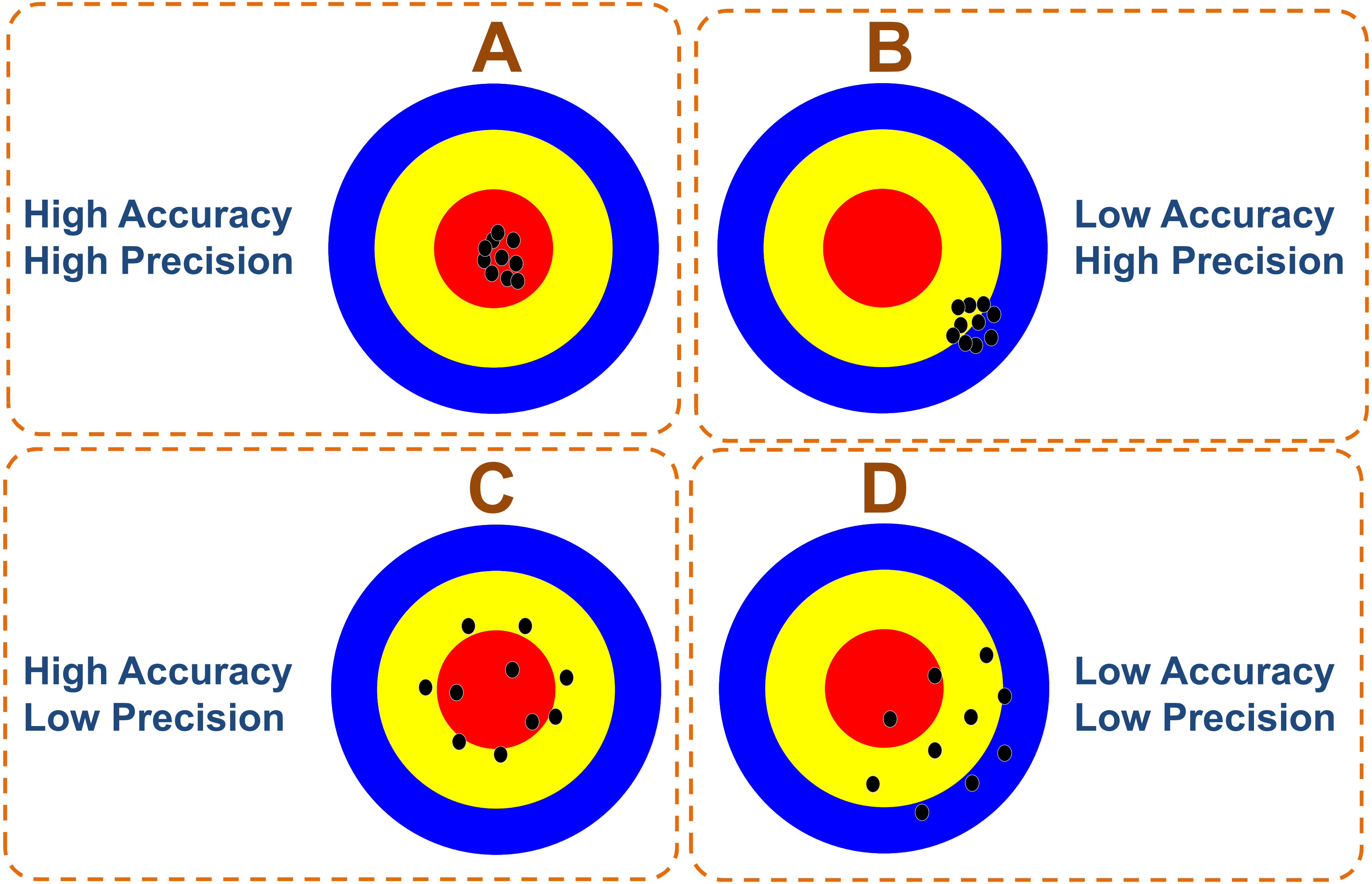
Accuracy Validity

Precision Reliability

Accuracy vs Prediction: Examples

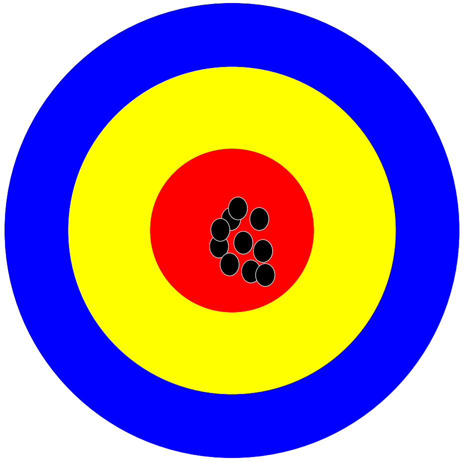


Accuracy vs. Precision



Accuracy vs. Precision

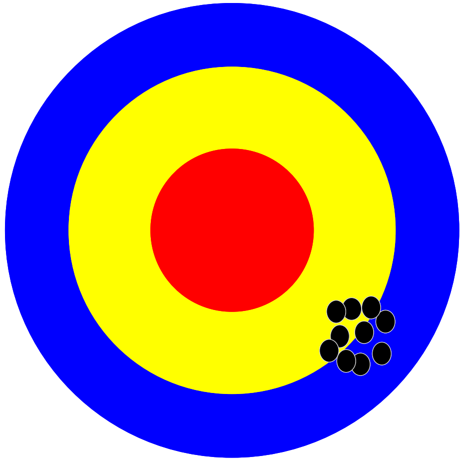
A



High Accuracy
High Precision

High Validity
High Reliability

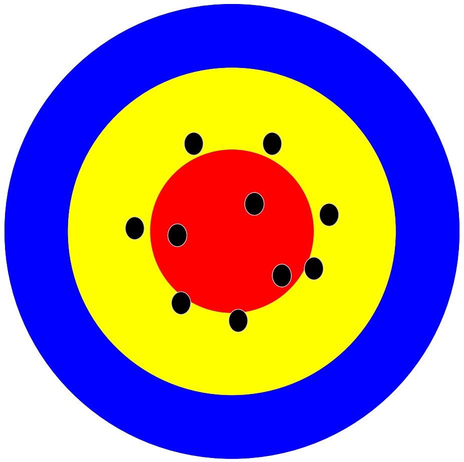
B



Low Accuracy
High Precision

Low Validity
High Reliability

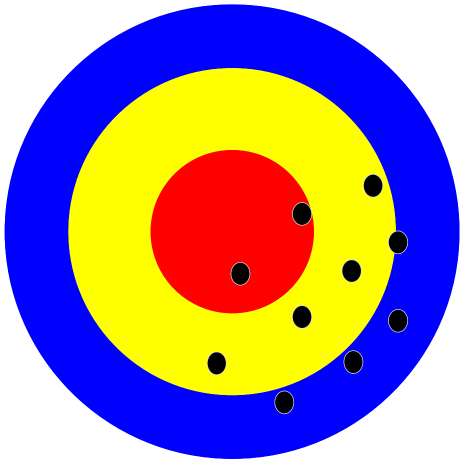
C



High Accuracy
Low Precision

High Validity
Low Reliability

D

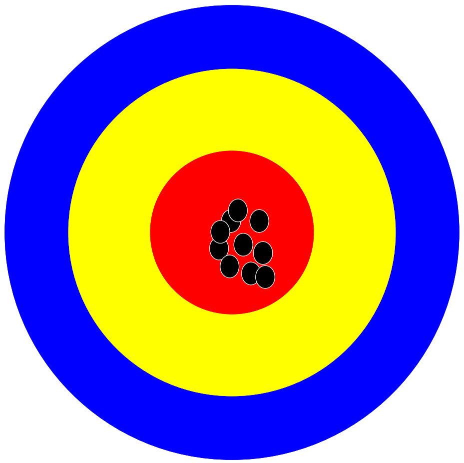


Low Accuracy
Low Precision

Low Validity
Low Reliability

Accuracy vs. Precision

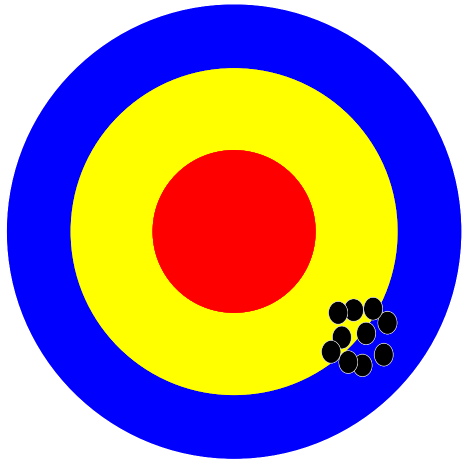
A



High Accuracy
High Precision

High Validity
High Reliability

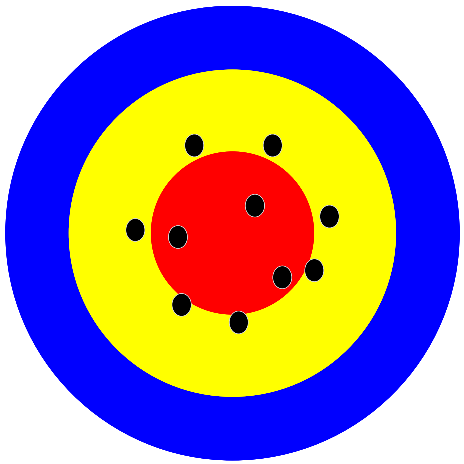
B



Low Accuracy
High Precision

Low Validity
High Reliability

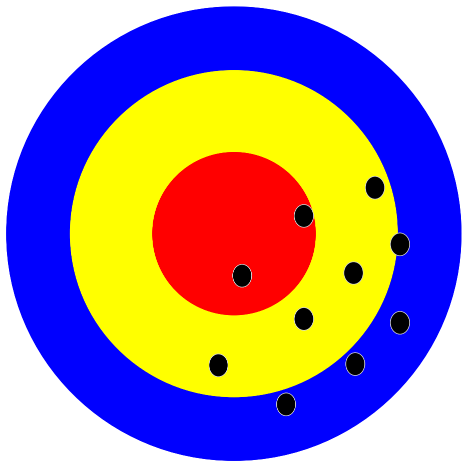
C



High Accuracy
Low Precision

High Validity
Low Reliability

D



Low Accuracy
Low Precision

Low Validity
Low Reliability

Please remember:

		True Class	
		Positive	Negative
Predicted Class	Positive	True Positive Count (TP)	False Positive Count (FP)
	Negative	False Negative Count (FN)	True Negative Count (TN)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$True\ Positive\ Rate = \frac{TP}{TP + FN}$$

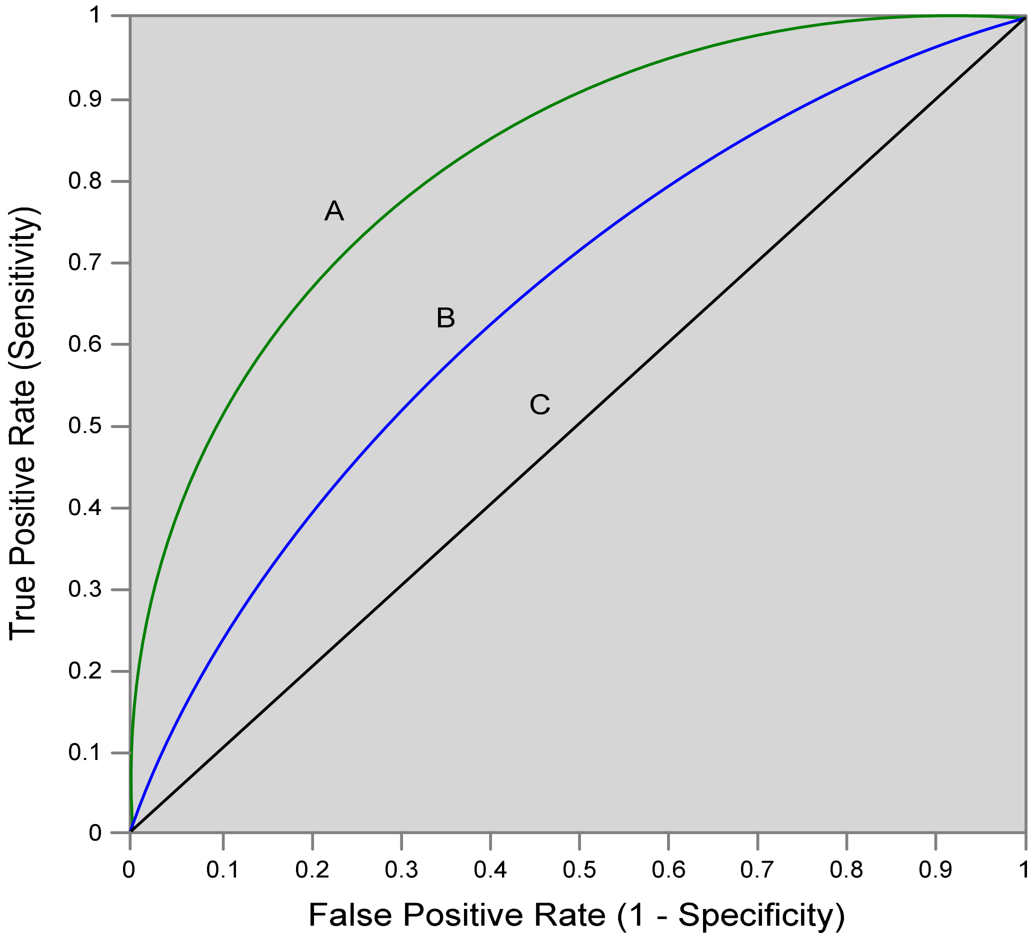
$$True\ Negative\ Rate = \frac{TN}{TN + FP}$$

$$Precision = \frac{TP}{TP + FP} \qquad Recall = \frac{TP}{TP + FN}$$

Turban, E., Sharda, R., Delen, D. & Efraim, T. 2007. Decision support and business intelligence systems, Pearson Education.

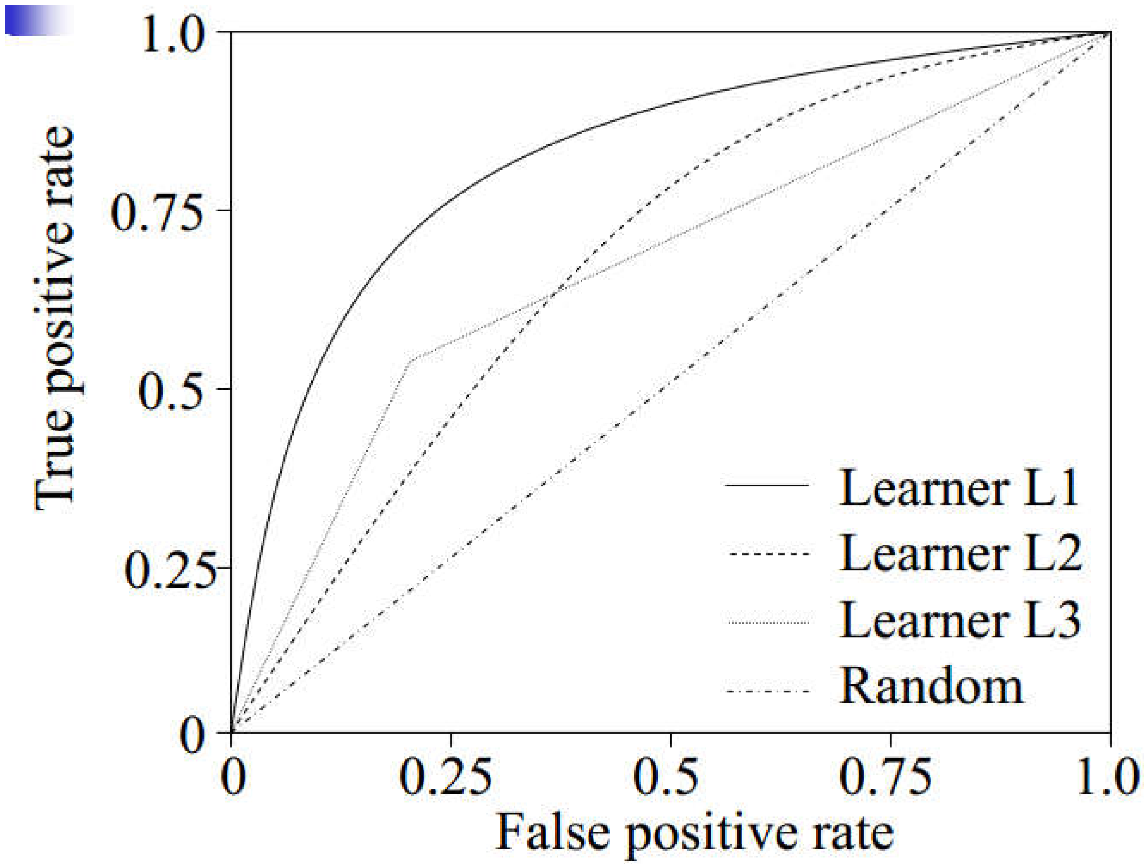
75

Again the ROC Curve



Bradley, A. P. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition,* 30, (7), 1145-1159.

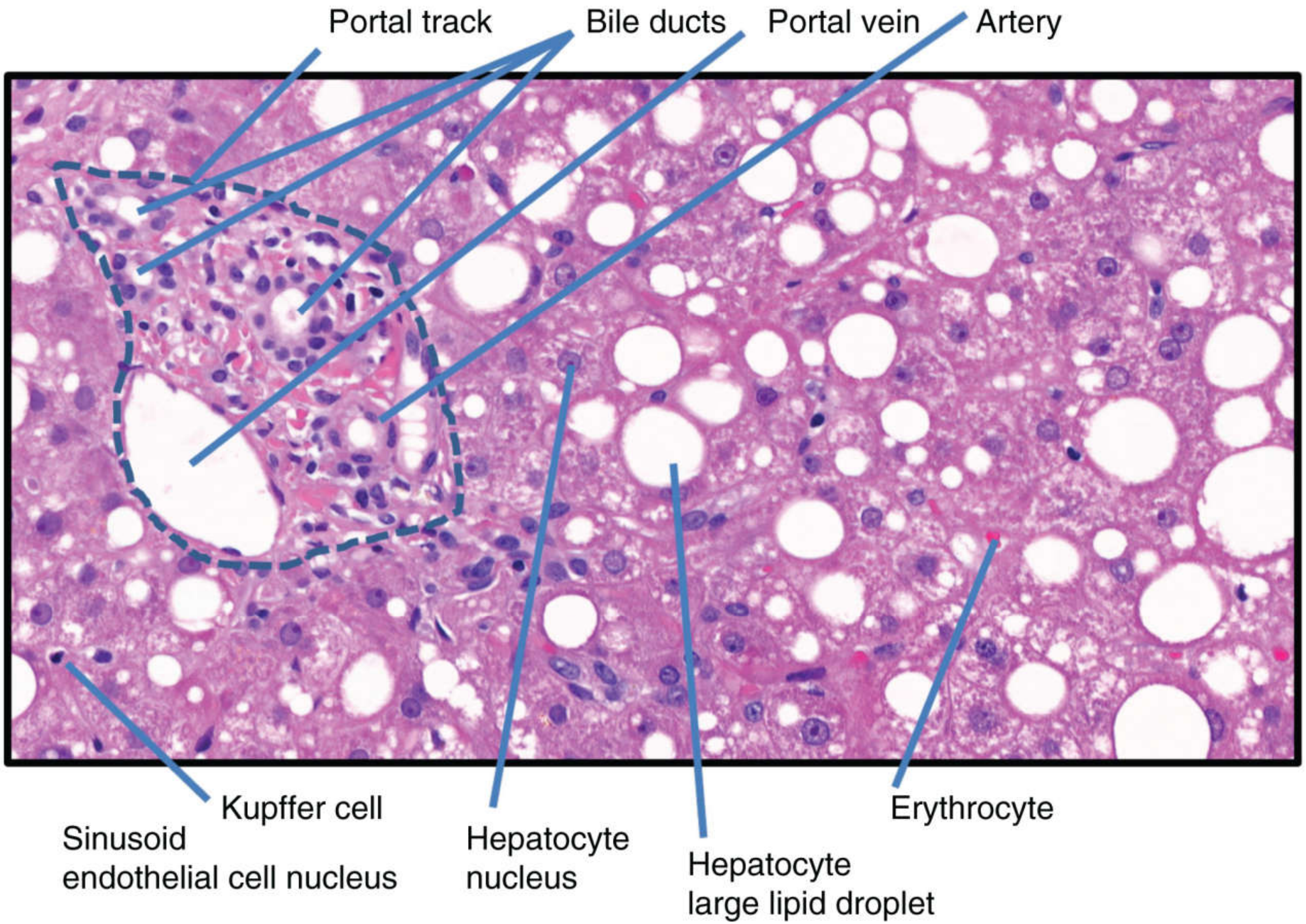
Source: Turban et al. (2011), Decision Support and Business Intelligence Systems 76



For a detailed explanation refer to: Fawcett, T. 2006. An introduction to ROC analysis. Pattern recognition letters, 27, (8), 861-874.

08 #KandinksyPatterns – our "Swiss-Knife" for the study of explainable AI

Motivation: Typical Task of a human Pathologist …



Andreas Holzinger, Georg Langs, Helmut Denk, Kurt Zatloukal & Heimo Mueller 2019. Causability and Explainability of AI in Medicine. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, doi:10.1002/widm.1312.

The medical report is legally binding – not the image

angelegt am 06.05.2006/20:26
geschr. von
gedruckt am 17.11.2006/08:24
Anfo: NCHIN

Radiologischer Befund

Kurzanamnese: St.p. SHT

Fragestellung: -

Untersuchung: **Thorax eine Ebene liegend**

SB

Bewegungsartefakte. Zustand nach Schädelhirntrauma.

Das Cor in der Größennorm, keine akuten Stauungszeichen.
Fragliches Infiltrat parahilär li. im UF, RW-Erguss li.

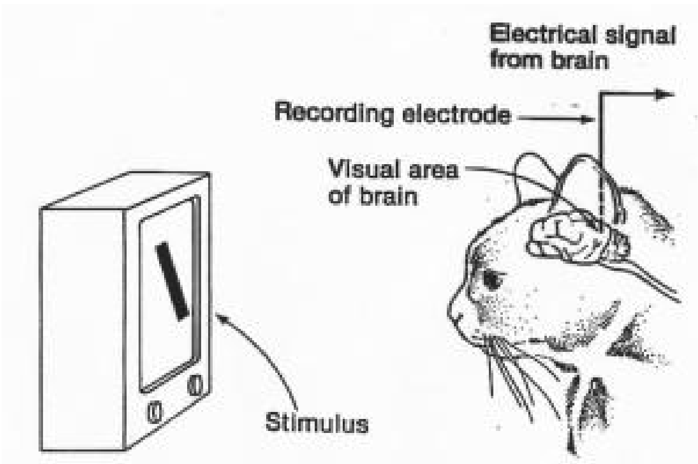
Zustand nach Anlage eines ET, die Spitze ca. 5cm cranial der Bifurkation, lieg. MS, orthotop positioniert. ZVK über re., die Spitze in Proj. auf die VCS. Kein Hinweis auf Pneumothorax. Der re. Rezessus frei.

Mit kollegialen Grüßen

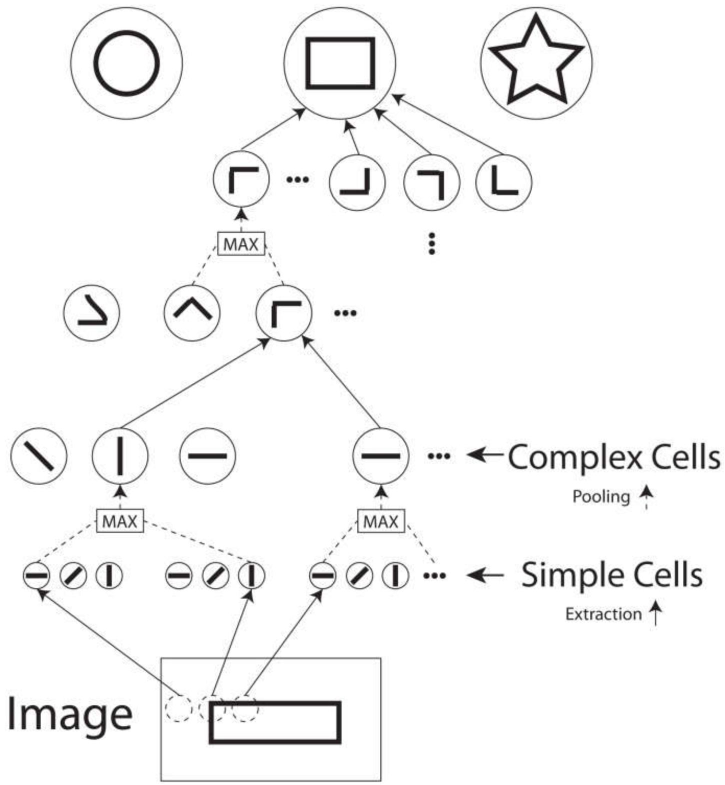
**** Elektronische Freigabe durch am 09.05.2006 ****

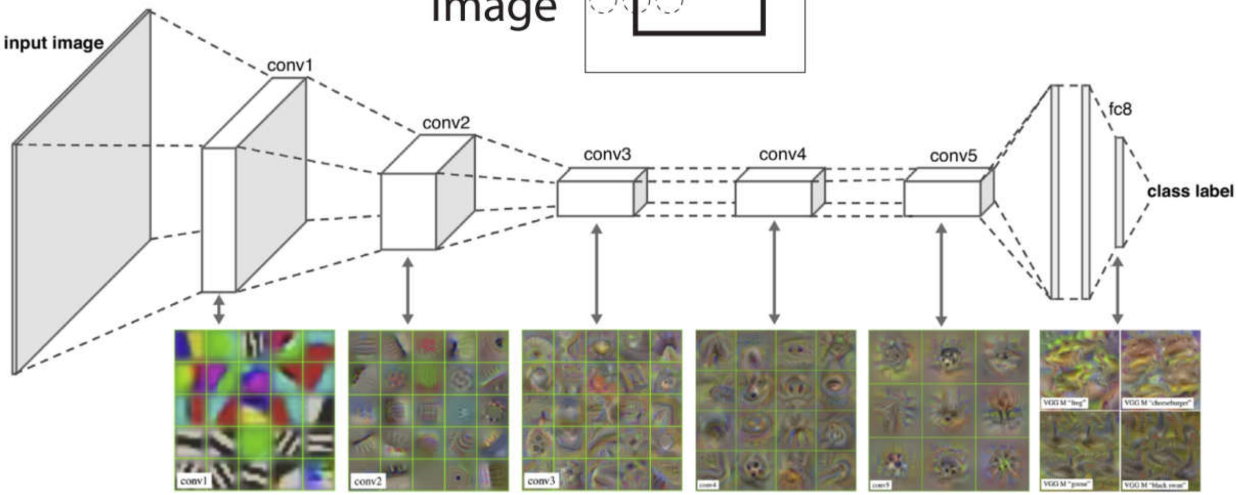
Holzinger, A., Geierhofer, R. & Errath, M. 2007. Semantische Informationsextraktion in medizinischen Informationssystemen. *Informatik Spektrum, 30, (2), 69-78.*

Hubel & Wiesel (1962): Our world is compositional

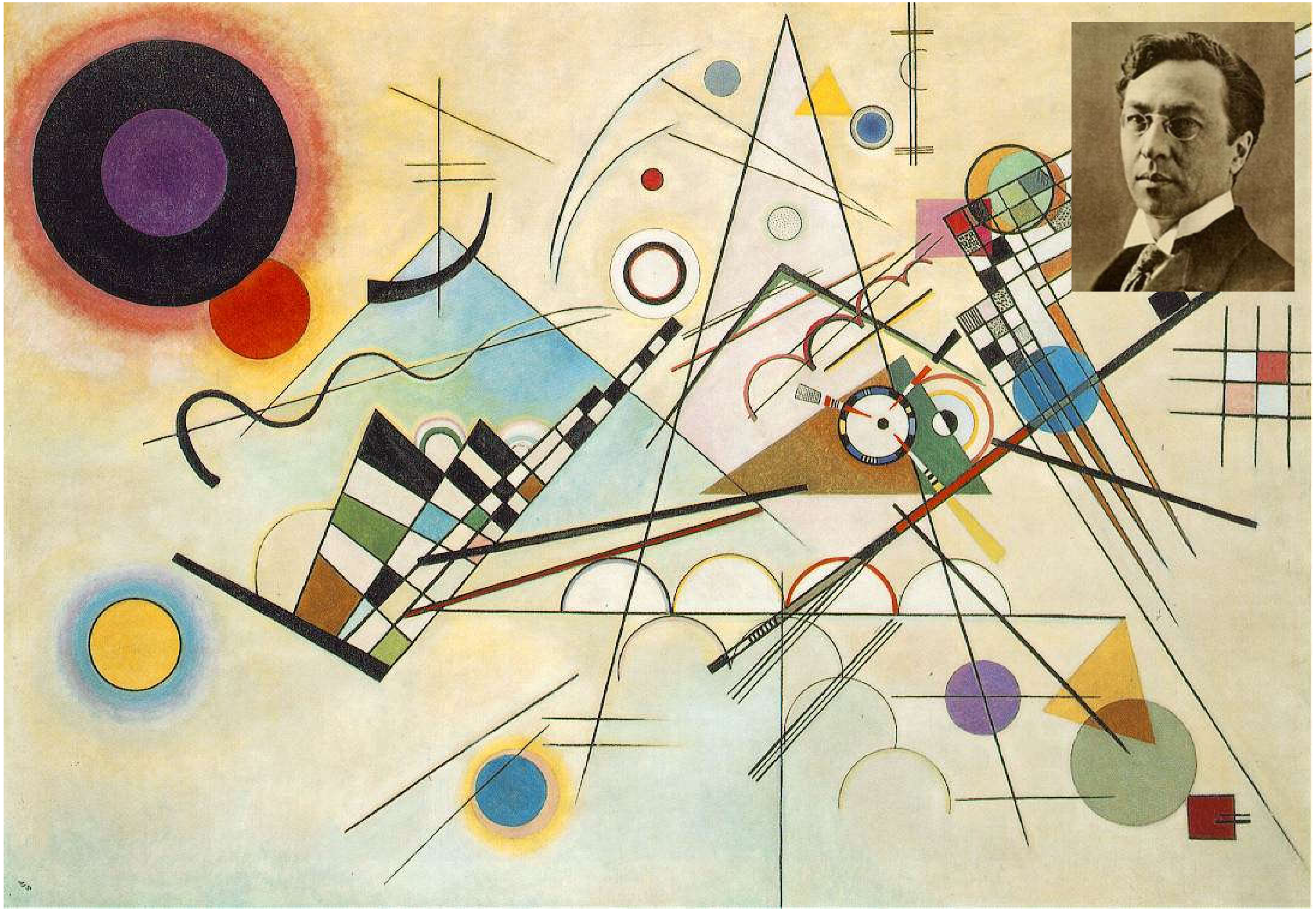


David H. Hubel & Torsten N. Wiesel 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. The Journal of Physiology, 160, (1), 106-154, doi:10.1113/jphysiol.1962.sp006837.



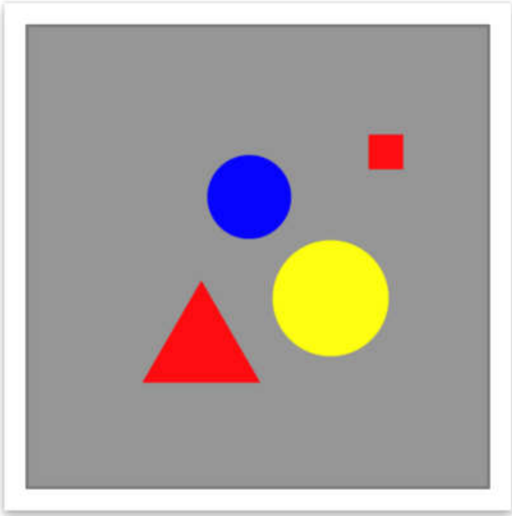


Wassily Kandinsky (1866 – 1944)



Komposition VIII, 1923, Solomon R. Guggenheim Museum, New York.Source: https://de.wikipedia.org/wiki/Wassily_Kandinsky
This images are in the public domain.

Definition 1: A Kandinsky Figure is …



- … a square image containing 1 to n geometric objects.
- Each object is characterized by its shape, color, size and position within this square.
- Objects do not overlap and are not cropped at the border.
- All objects must be easily recognizable and clearly distinguishable by a human observer.

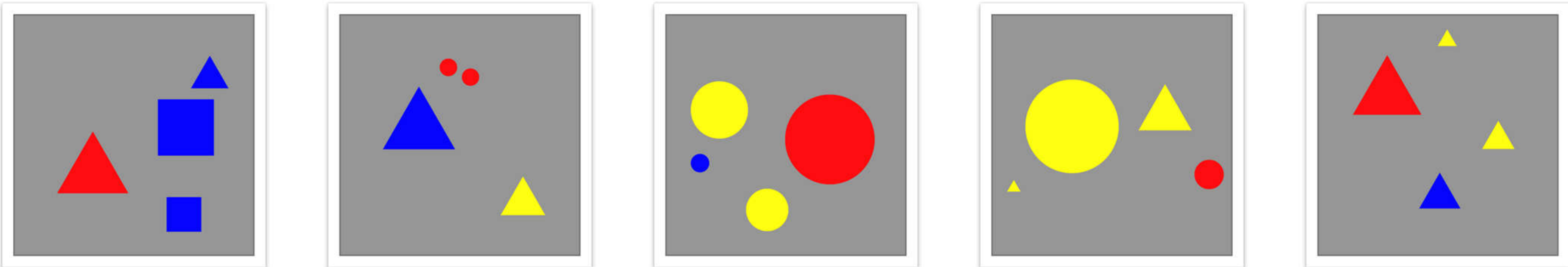
Definition 2 A statement $s(k)$

- about a Kandinsky Figure k is …
- either a mathematical function $s(k) \rightarrow B$; with B $(0,1)$
- or a $natural\ language\ statement$ which is true or false

- Remark: The evaluation of a natural language statement is always done in a specific context. In the followings examples we use **well known concepts from human perception** and linguistic theory.
- If $s(k)$ is given as an algorithm, it is essential that the function is a pure function, which is a computational analogue of a mathematical function.

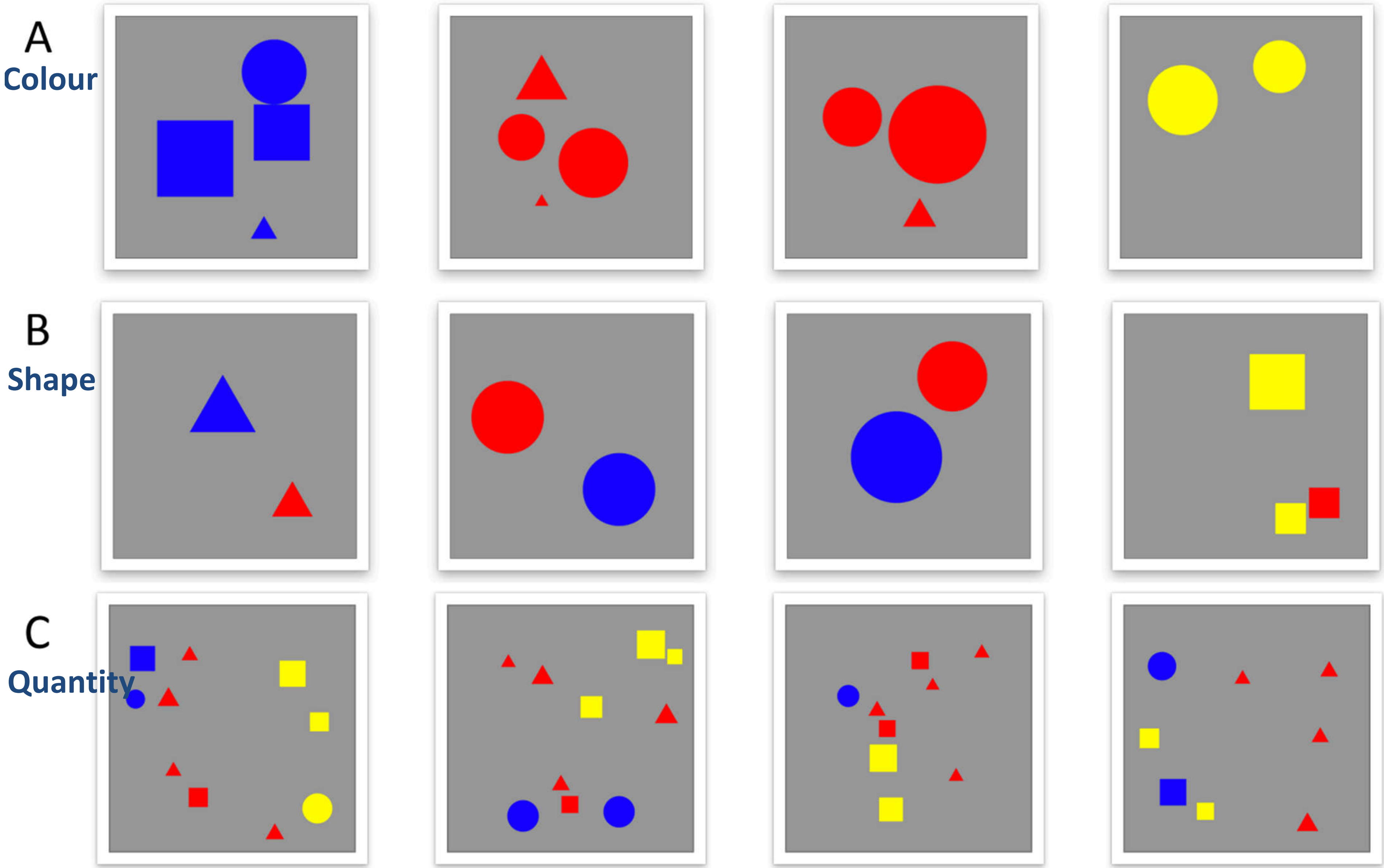
Definition 3 A Kandinsky Pattern K …

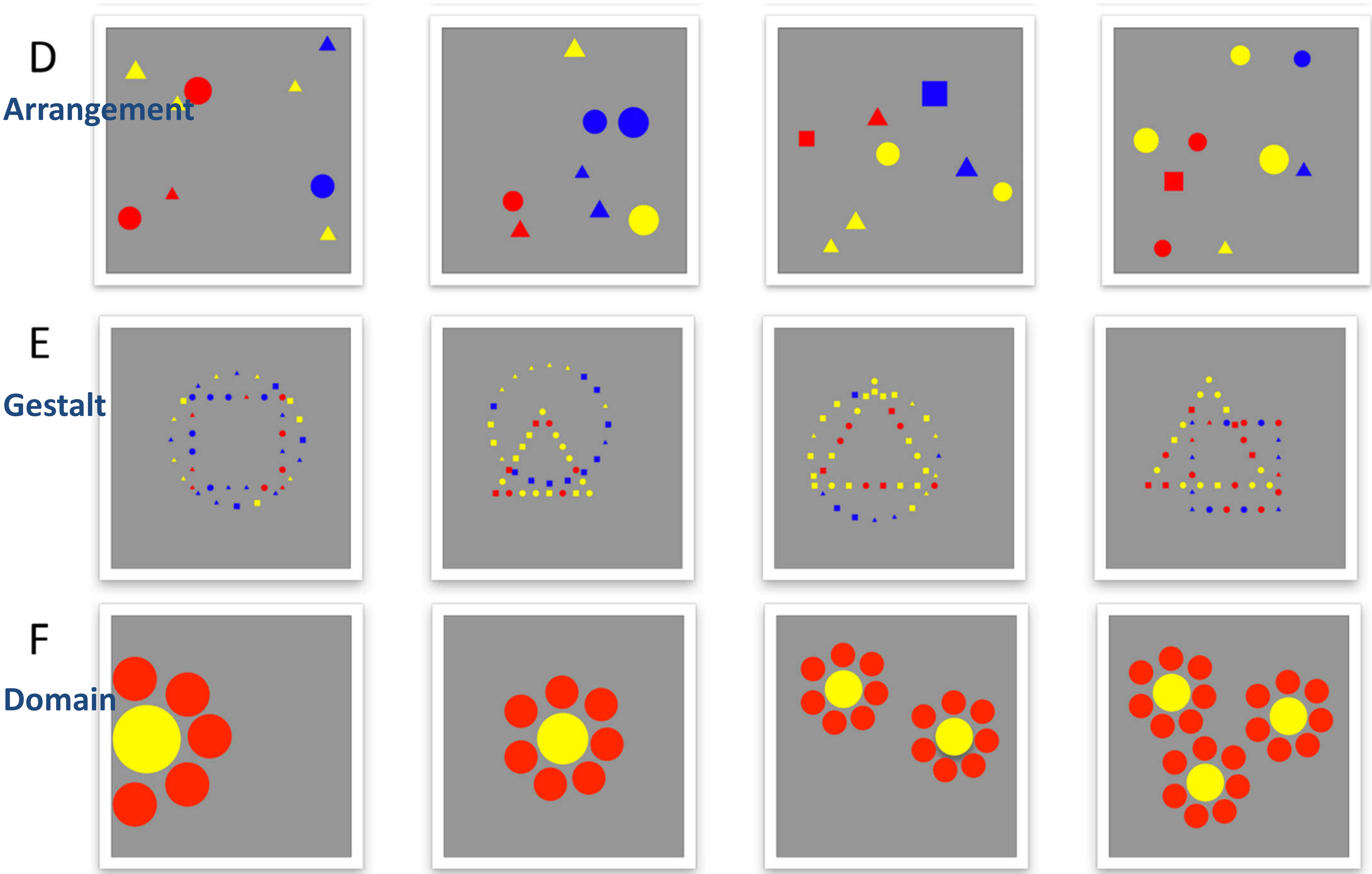
- … is defined as the subset of all possible Kandinsky Figures k with $s(k) \rightarrow 1$ or the natural language statement is true.
- $s(k)$ and a natural language statement are equivalent, if and only if the resulting Kandinsky Patterns contains the same Kandinsky Figures.
- $s(k)$ and the natural language statement are defined as the **Ground Truth** of a Kandinsky Pattern



"… the Kandinsky Figure has two pairs of objects with the same shape, in one pair the objects have the same color, in the other pair different colors, two pairs are always disjunct, i.e. they don't share a object …".

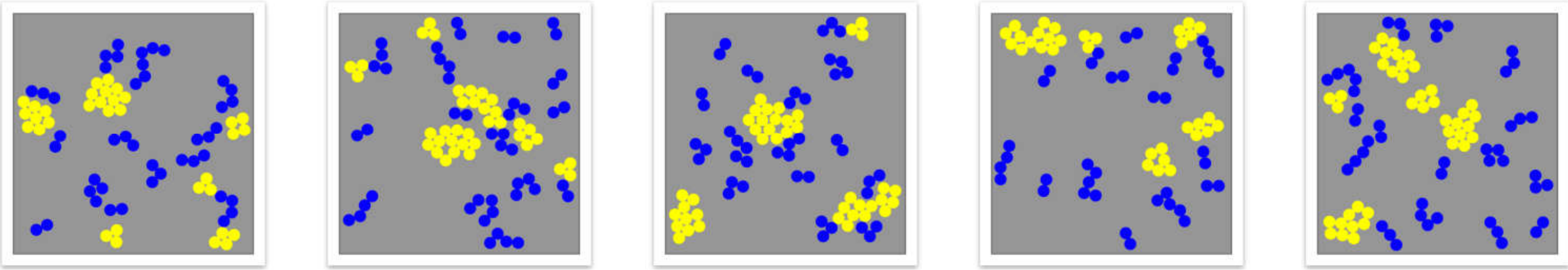
How do humans explain? They use concepts!



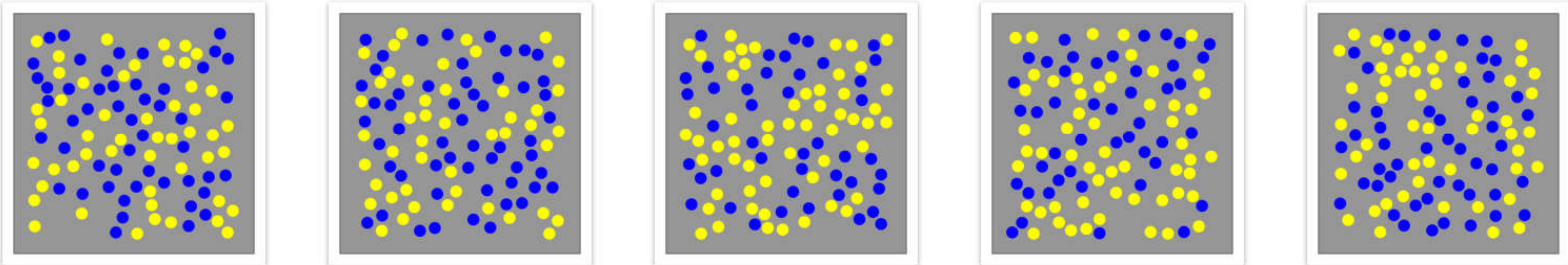

D
Arrangement
E
Gestalt
F
Domain

Domain Concept from Pathology

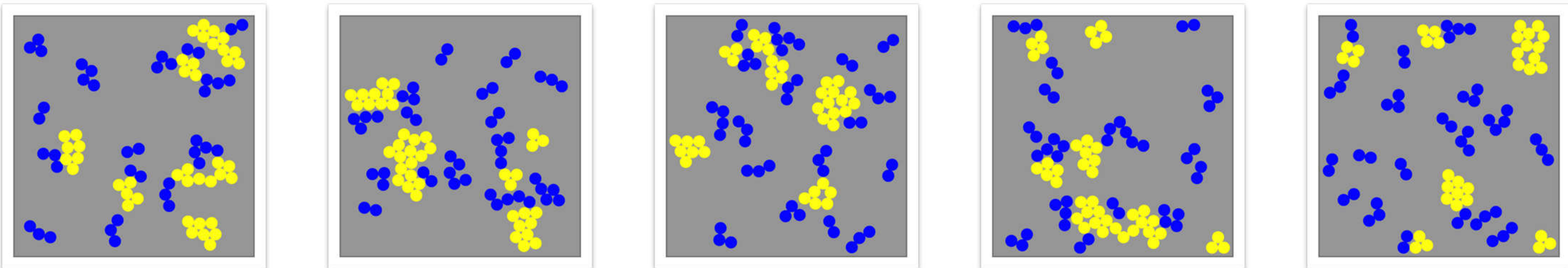
A) True



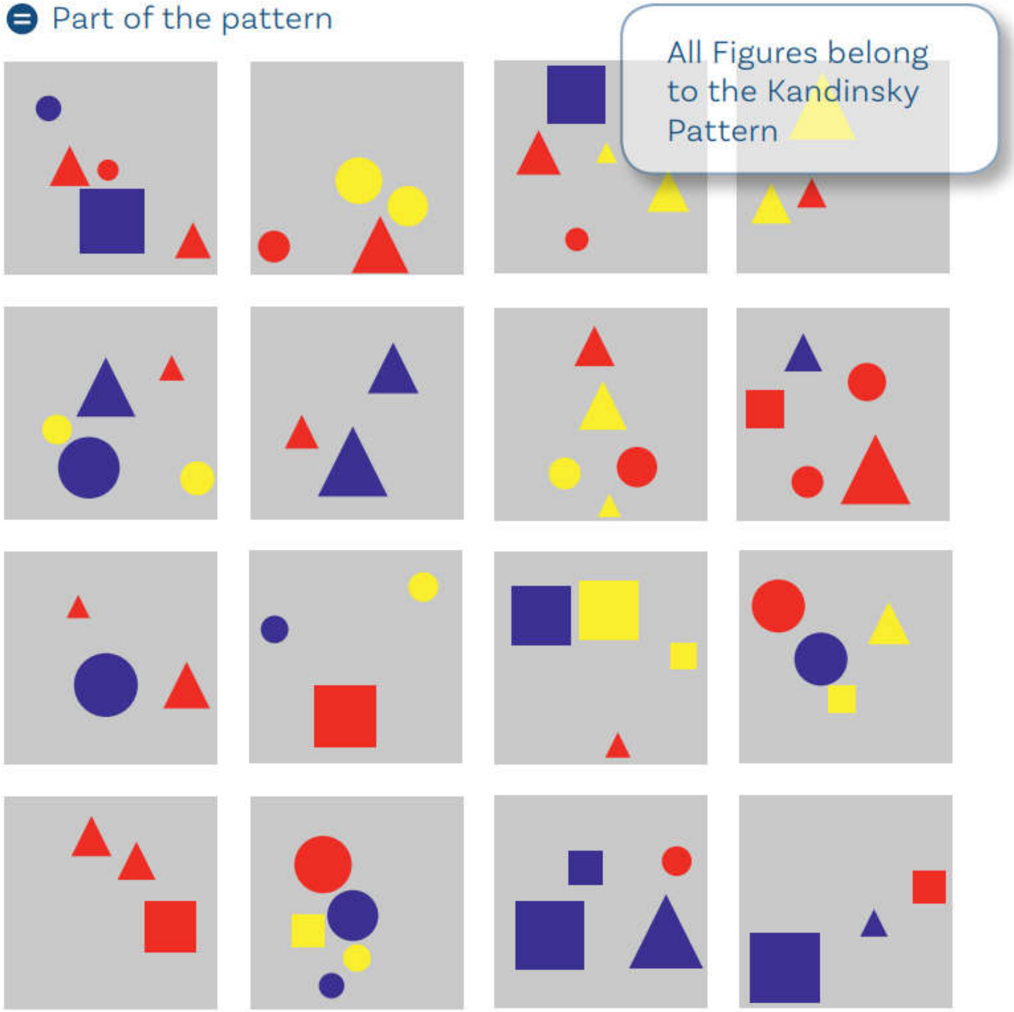
B) False



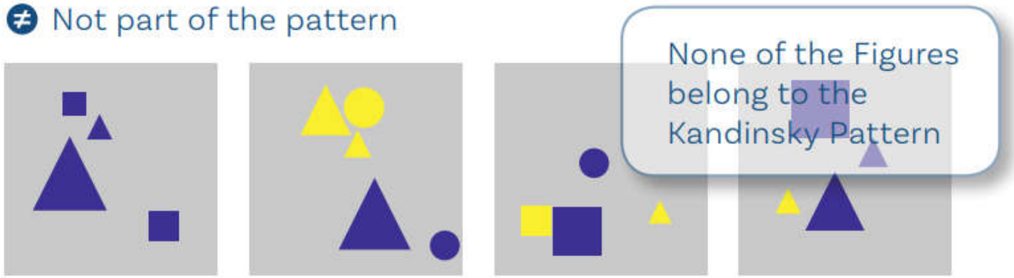
C) Counterfactual



Part of the pattern



Not part of the pattern



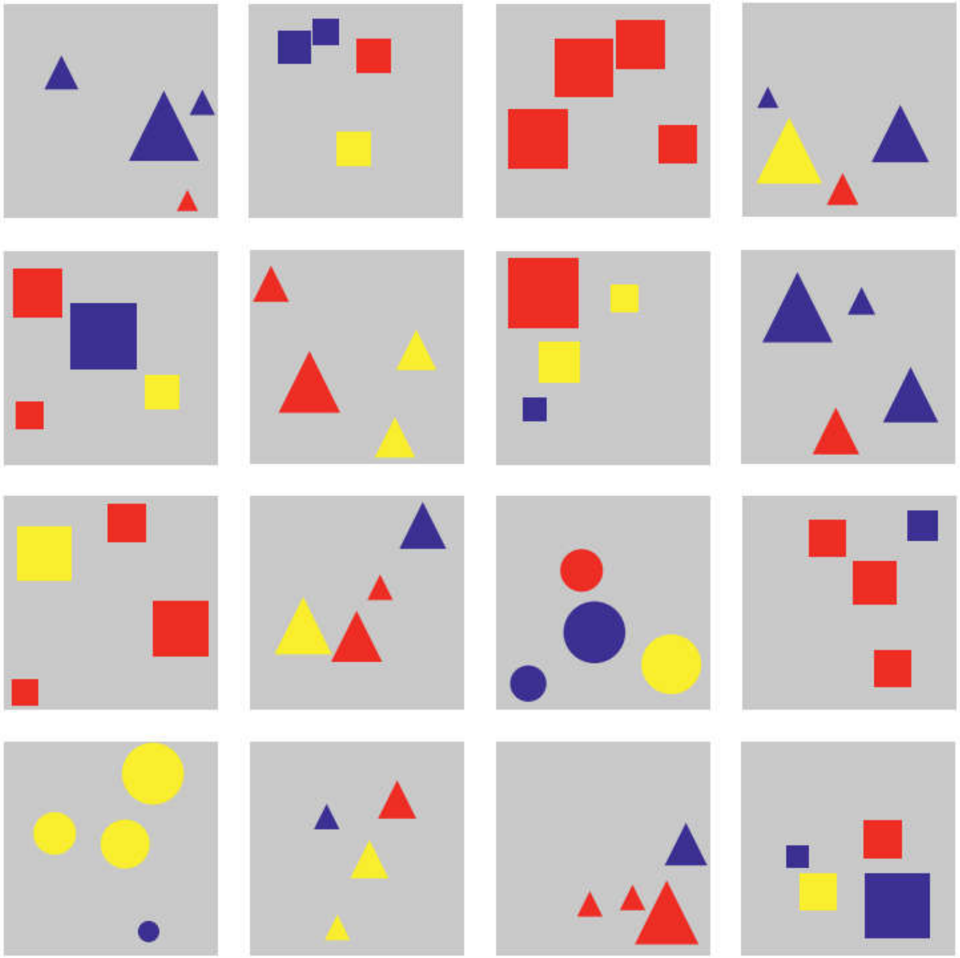
Hypothesis 1

It only contains circles and triangles.

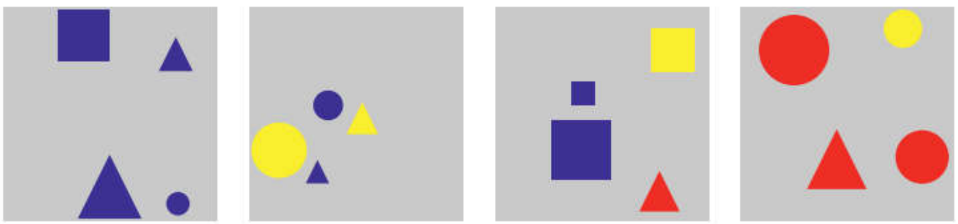
Hypothesis 2

It contains at least a red object.

Part of the pattern



Not part of the pattern



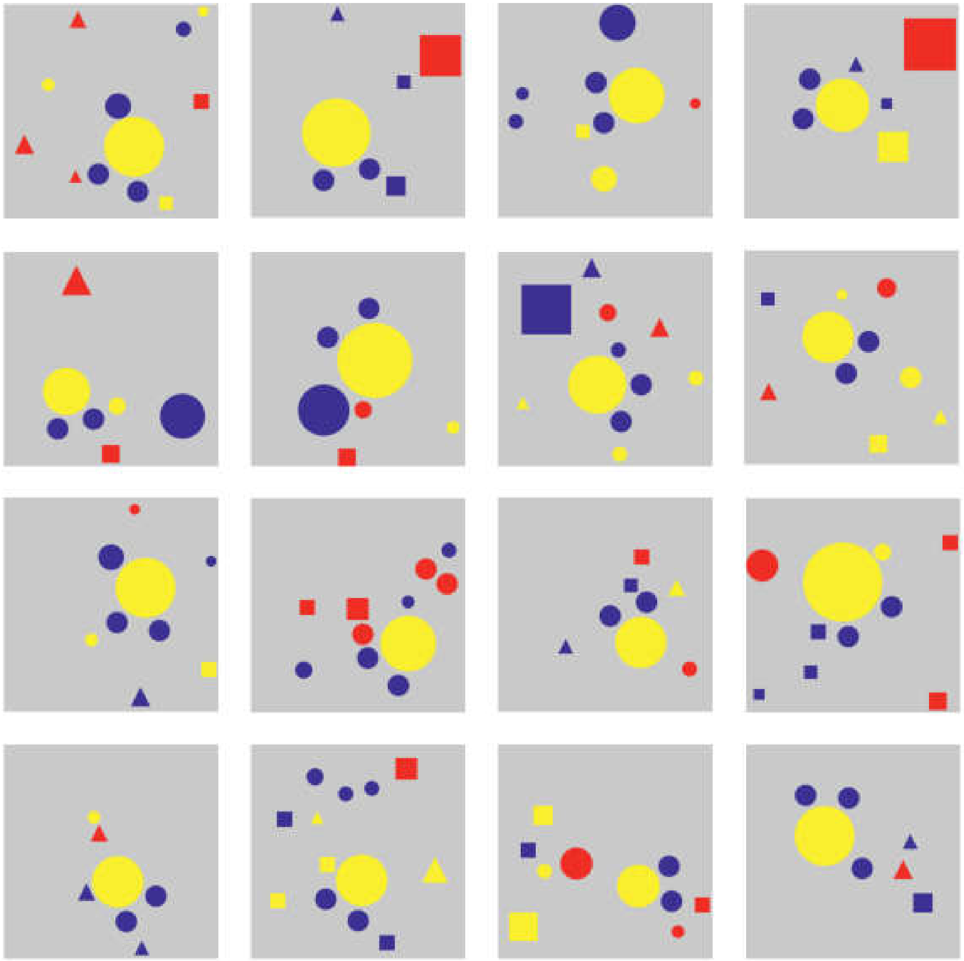
S2 Basic Pattern 2

Title: **All of Same Shape ->**

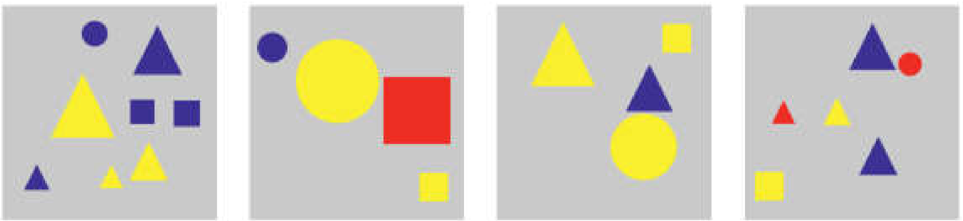
All objects have the same shape.

Hint: Don't be distracted by the colors

Part of the pattern



Not part of the pattern



S8 Basic Pattern 8

Title: **Mickey Mouse** ->

Every figure contains a pattern which is made out of a big yellow circle and two smaller blue ones and looks like a Mickey Mouse.

09 Sample Questions and Conclusion

Human-AI collaboration will sustainably influence the way science is done

How can I use visual representations of abstract data to amplify the acquisition of knowledge?

Hypothetico-Deductive Method

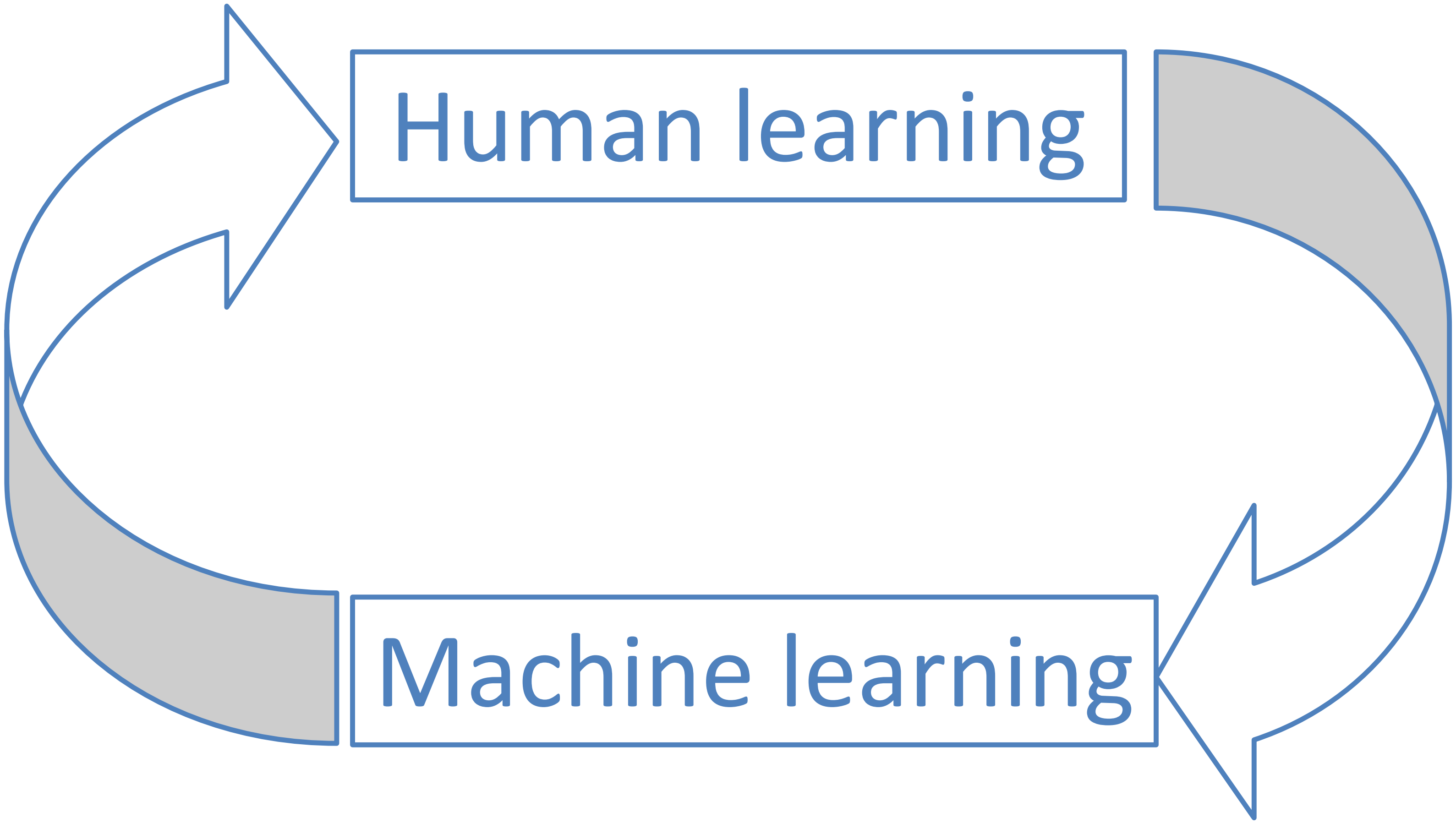
- 1) Given a set of (complex) data
- 2) Set a hypothesis
- 3) Extract information
- 4) Discover hidden knowledge
- 5) Support your previous set hypothesis
- **Machine intelligence + Human intelligence**
- = powerful methods for many sciences
- Application e.g. in many domains, e.g. medicine and health, education, psychology, industry, etc

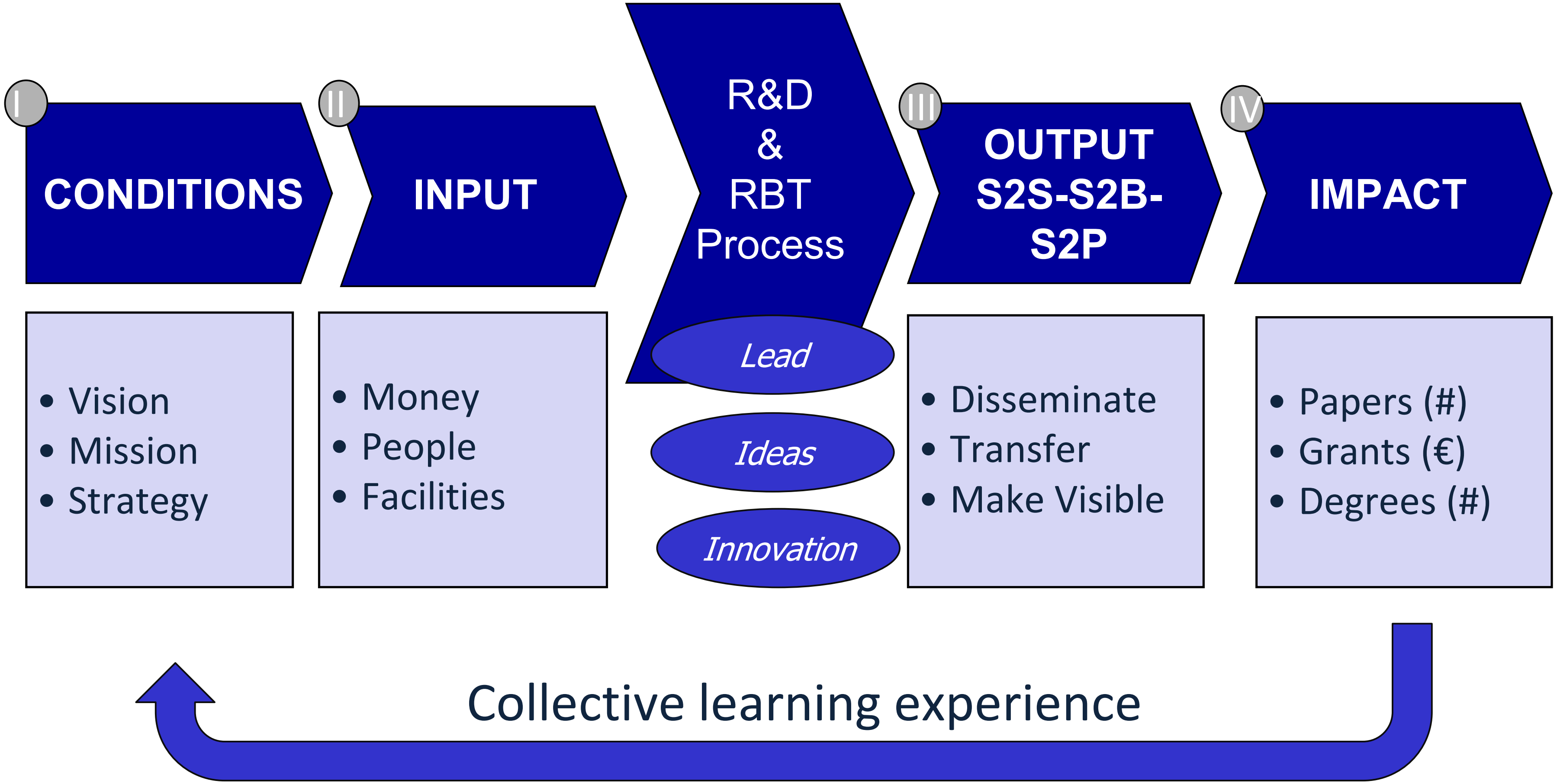
Some top-level grand questions (not ordered)

- How can we transfer learned representations to improve learning in other tasks/domains?
- Which learning algorithm should be used when?
- Can machine learning theories help to understand human learning and vice versa?
- Machine Learning vs. Human Learning: role of motivation, emotion, forgetting, … ?
- How can we use self-supervised learning with multiple sensory input?

https://ai100.stanford.edu/

Appendix


Human learning
Machine learning



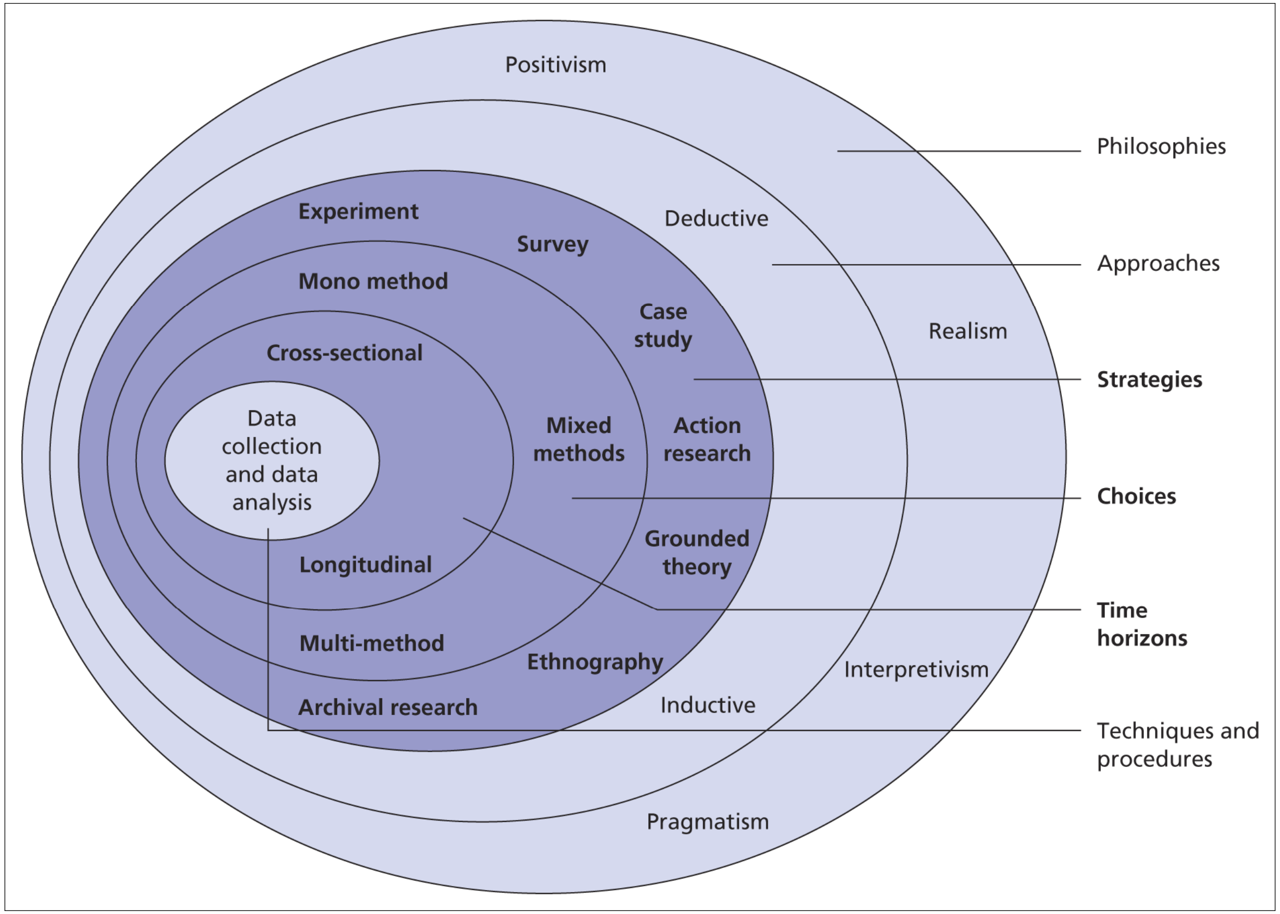
Holzinger, A. 2011. *Successful Management of Research and Development,* Norderstedt, BoD.

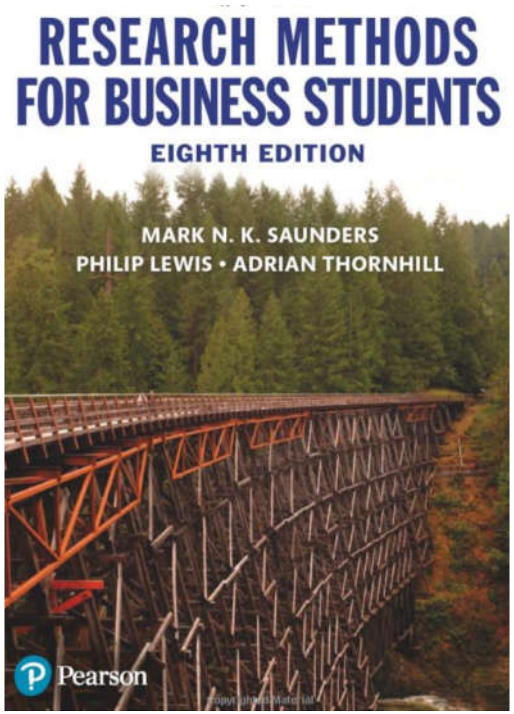
Essential Tools for Scientific Machine Learning and Scientific AI				Comparison of tools readily usable with differentiable programming (automatic differentiation) frameworks			
Subject | AD Frameworks	ADIFOR or TAF	ADOL-C	Stan	Julia (Zygote.jl, Tracker.jl, ForwardDiff.jl, etc.)	TensorFlow	PyTorch	Misc. other good packages
Language	Fortran	C++	Misc.	Julia	Python, Swift, Julia, etc.	Python	
Neural Networks	neural-fortran	OpenNN	None	Flux.jl	Built-in	Built-in	ADIFOR
Neural Differential Equations	Sundials (ODE+DAE)	Sundials (ODE+DAE)	Sundials (ODE+DAE)	DifferentialEquations.jl / DiffEqFlux.jl (ODE, SDE, DDE, DAE, hybrid, (S)PDE)	DifferentialEquations.jl (through Tensorflow.jl)	torchdiffeq (non-stiff ODEs)	PyMC3 (Python)
	FATODE	PETSc TS	Built-in (non-stiff ODE)	Sundials.jl (ODE through DiffEqFlux.jl)		diffeqpy	SMT (Python)
Probabilistic Programming	None	CPProb	Built-In	Gen.jl	Edward	Pyro	sensitivity (R)
				Turing.jl	PyMC4	pyprob	ColPACK (Fortran)
Sparsity Detection	Built-in (TAF)	Built-in	None	SparsityDetection.jl	None	None	Dakota
Sparse Differentiation	Built-in (TAF)	Built-in	None	SparseDiffTools.jl	None	None	PSUDAE
GPU Support	CUDA	CUDA	OpenCL	CUDANative.jl + CuArrays.jl	Built-in	Built-in	Mondrian
Distributed Dense Linear Algebra	ScaLAPACK	Elemental	None	Elemental.jl	Built-in	torch.distributed (no factorizations)	SimLab (MATLAB)
				DistributedArrays.jl		Elemental	Halide
Distributed Sparse Linear Algebra	ScaLAPACK	PETSc	None	Elemental.jl	Built-in (no factorizations)	Elemental	
	PARASOL	Trilinos		PETSc.jl	None	petsc4py	
		Elemental			None		
Structured Linear Algebra	SPARSEKIT	None	None	(Block)BandedMatrices.jl	Some built-in	None	
Surrogate Modeling	None	MUQ	None	Surrogates.jl	None	pySOT	
Global Sensitivity Analysis	None	None	None	DifferentialEquations.jl / DiffEqSensitivity.jl	None	SALib	
Uncertainty Quantification	None	MUQ	None	DifferentialEquations.jl / DiffEqUncertainty.jl	None	uncertainpy	
Direct Distributed Parallelism	MPI	MPI	MPI	MPI.jl	Built-in	torch.distributed	
PDE Discretizations	C++ Libraries w/ Fortran interfaces (hypre)	deal.ii	None	ApproxFun.jl		FEniCS	
		SAMRAI		DiffEqOperators.jl		Firedrake	
		hypre		JuAFEM / JuliaFEM		Dedalus	

http://www.stochasticlifestyle.com/the-essential-tools-of-scientific-machine-learning-scientific-ml

For more details, see http://www.stochasticlifestyle.com/a-comparison-between-differential-equation-solver-suites-in-matlab-r-julia-python-c-and-fortran

Scale	None	Poor	Fair	Excellent
Explanation	No automatic differentiation compatible library exists. Suggestion for a library to wrap.	Functionality exists, but is feature-incomplete or AD compatibility is incomplete. If no AD support, then AD support can easily be added since the library already defines adjoints.	The basic features exist, but has some major features missing or are not AD-compatible.	Has all of the main features and is fully compatible with the automatic differentiation tooling.





https://www.birmingham.ac.uk/schools/business/staff/profile.aspx?ReferenceId=104618&Name=professor-mark-nk-saunders

e-Society

Science

What is information?
What is intelligence?
What is computable?

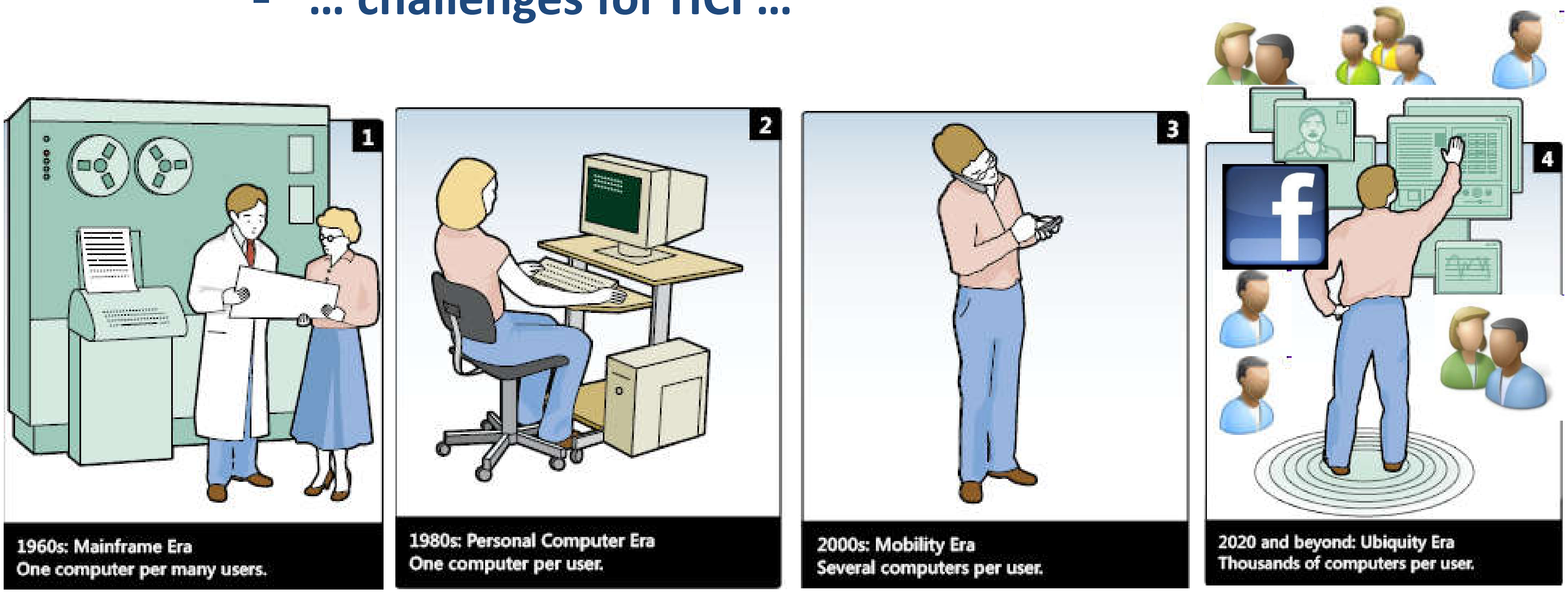
Engineering

(How) can we build
these systems
simply?

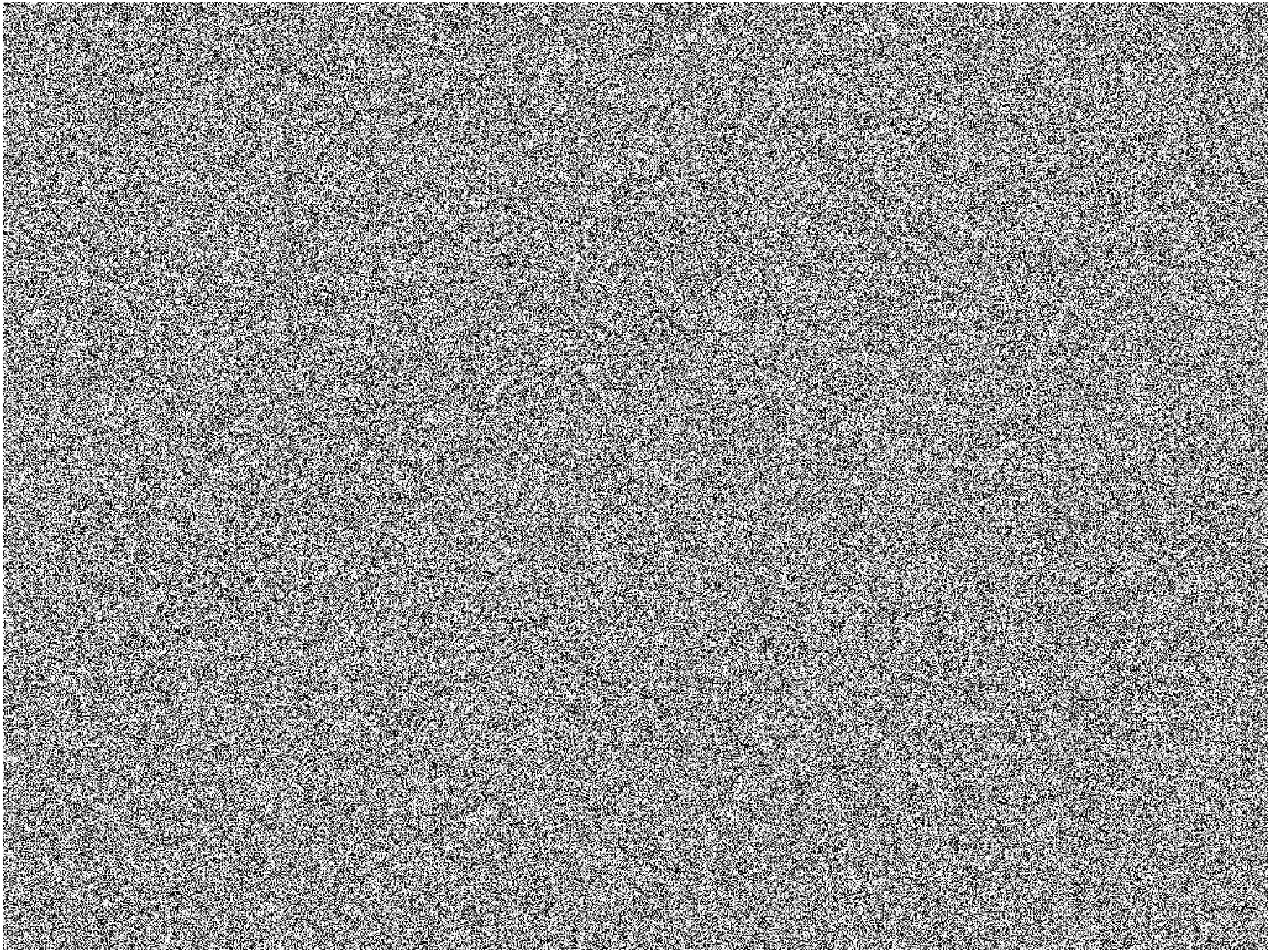
Kleine Einheiten			Große Einheiten		
10^{-3}	milli	m	10^{3}	kilo	k
10^{-6}	micro	µ	10^{6}	mega	M
10^{-9}	nano	n	10^{9}	giga	G
10^{-12}	pico	p	10^{12}	tera	T
10^{-15}	femto	f	10^{15}	peta	P
10^{-18}	atto	a	10^{18}	exa	E
10^{-21}	zepto	z	10^{21}	zetta	Z
10^{-24}	yocto	y	10^{24}	yotta	Y

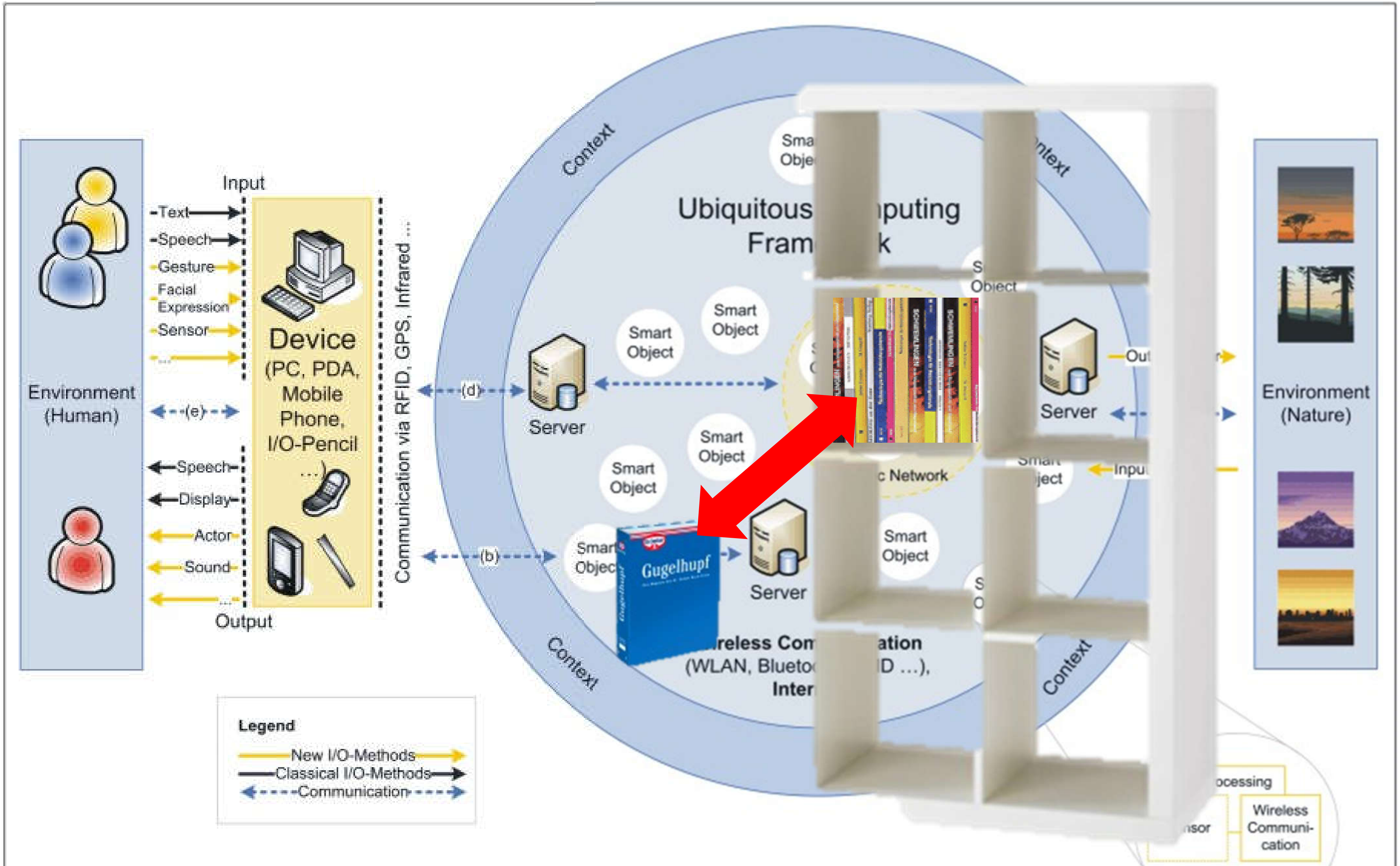
Please note: Computers change constantly …

- Old dream of mankind: using technology to augment human capabilities for structuring, retrieving and managing information and decision support:
- **… challenges for HCI …**



Harper, Rodden, Rogers, Sellen (2008)





Andreas Holzinger, Alexander Nischelwitzer, Silvia Friedl & Bo Hu 2010. Towards life long learning: three models for ubiquitous applications. Wireless Communications and Mobile Computing, 10, (10), 1350-1365, doi:10.1002/wcm.715.

Let the user interactively manipulate the data

- **Focus Selection** = via direct manipulation and selection tools, e.g. multi-touch (in data space a n-dim location might be indicated); see a recent work by Randy Goebel
- **Attention Routing** = anomaly detection, draws people's attention to interesting areas to start their analyses;
- **Extent Selection** = specifying extents for an interaction, e.g. via a vector of values (a range for each data dimension or a set of constraints;
- **Interaction type selection** = e.g. a pair of menus: one to select the space, and the other to specify the general class of the interaction;
- **Interaction level selection** = e.g. the magnitude of scaling that will occur at the focal point (via a slider, along with a reset button);

Time is essential !!

DOI:10.1145/2500886

Timing analysis for hard real-time systems.

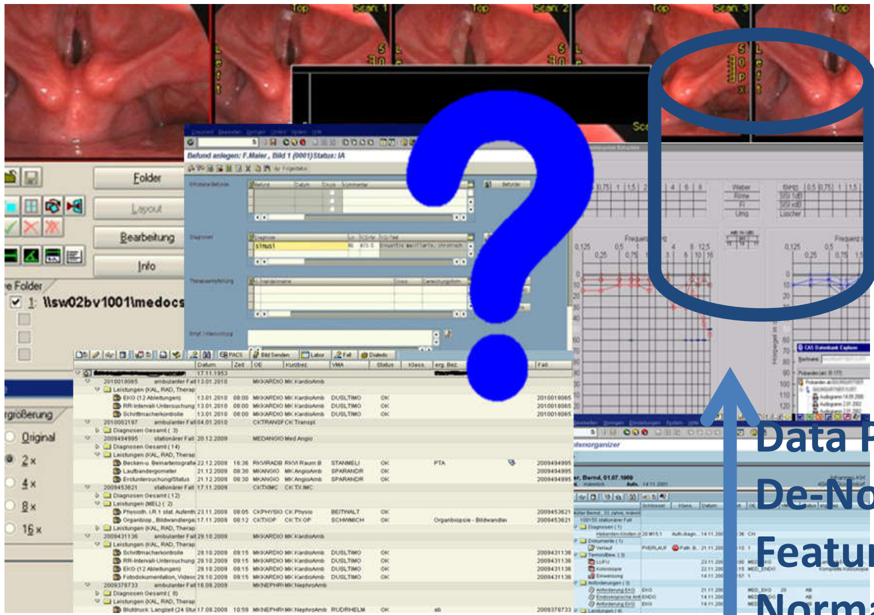
BY REINHARD WILHELM AND DANIEL GRUND

Computation Takes Time, But How Much?

Wilhelm, R. & Grund, D. 2014. Computation takes time, but how much? *Communications of the ACM, 57, (2), 94-103.*

Selected
Target Data

Visualization,
Validation
Etc.



Data Preprocessing
De-Noising
Feature extraction
Normalization etc.

Classification
Clustering etc.

Data Integration, Data Fusion
Privacy, Security, Safety
Data Protection
Anonymization-> Open Data

INTERACTION ?

Holzinger, A. & Zupan, M. 2013. KNODWAT: A scientific framework application for testing knowledge discovery methods for the biomedical domain. *BMC Bioinformatics, 14, (1), 191.*

Our central hypothesis:
Information bridges this gap

Simonic, K.-M. & Holzinger, A. (2010) Zur Bedeutung von Information in der Medizin. *OCG Journal, 35,* ***1, 8.***