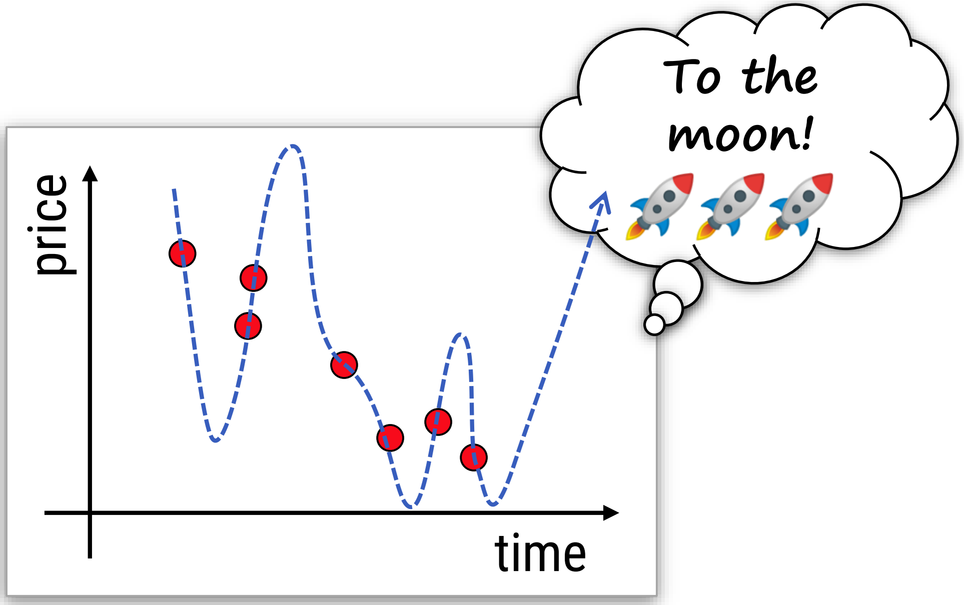
Modelling 2

STATISTICAL DATA MODELLING



Chapter 1

Knowledge & Uncertainty

Michael Wand · Institut für Informatik, JGU Mainz · michael.wand@uni-mainz.de

Introduction

What is the topic of this lecture?

Topic of This Lecture

Statistical Data Modeling

- Extracting knowledge from data
- Algorithmic inductive reasoning
- Statistical machine learning

Focus

- Self-organizing ML systems
- Algorithmic: Deep Networks
- “Emergent” structure
- Complex system modeling techniques

Motivation: “Artificial Intelligence”

Renewed interest

- Every 20 years?
- There is no AI (yet)

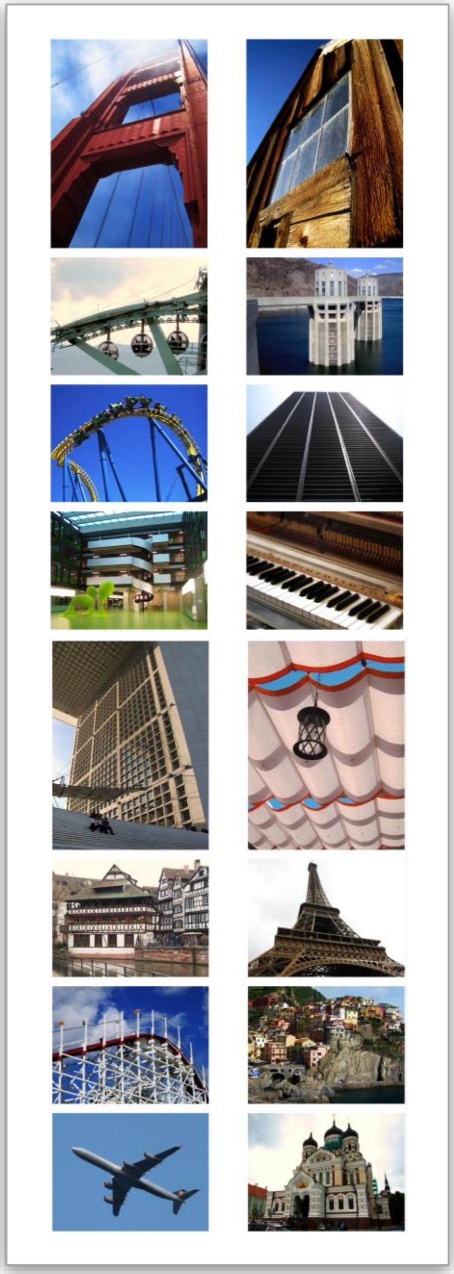
Research

- What is intelligence?
- Old question, unresolved
  - Philosophy, Physics, Biology
  - If we want to rebuild it, we have to find out
- Algorithmic formalization of intelligence

"Artificial Intelligence"

Statistical Data Modeling

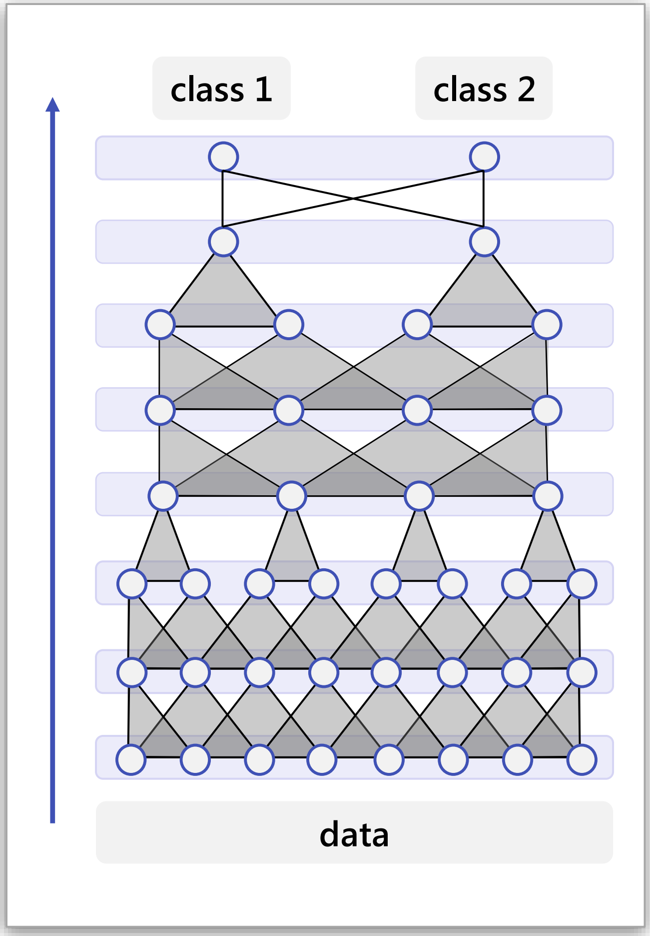
- Perspective: Intelligence
  - Make sense of the data around us
  - Uncertainty leads to statistics
  - Machine learning: algorithmic statistics



~~Artificial~~ Intelligence

Statistical Data Modeling

- Perspective: Intelligence
  - Make sense of the data around us
  - Uncertainty leads to statistics
  - Machine learning: algorithmic statistics
- Tool of the day: Deep Networks
  - Remarkable performance
  - Remarkably simple
  - **Why** do they work?



~~Modelling 2~~

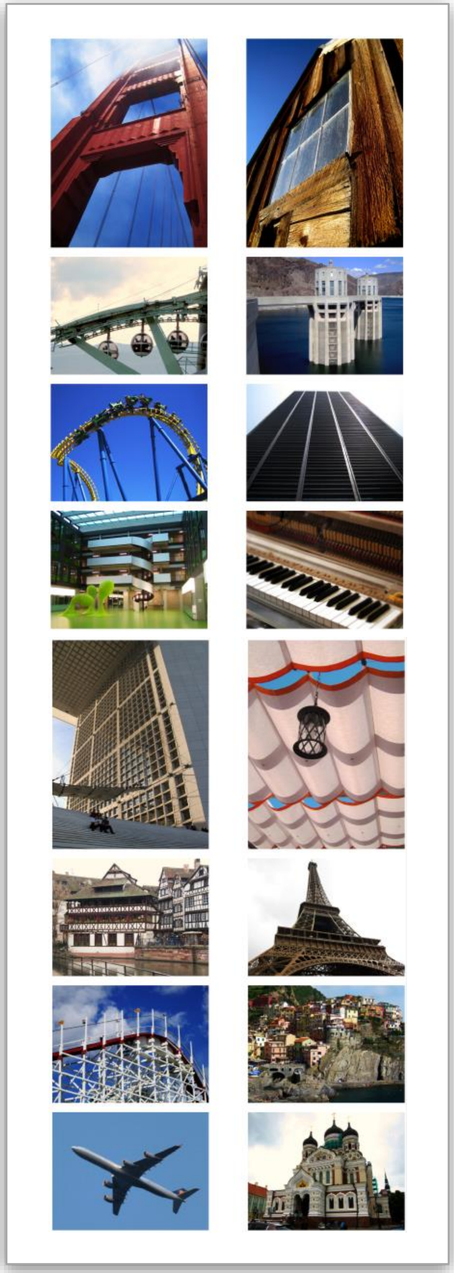
~~Statistical Data Modelling~~

Theoretical Deep Learning

~~Artificial~~ Intelligence

Statistical Data Modeling

- Perspective: Modeling
  - Model structure in data
  - Data can be cognitive system itself
  - Reverse engineering of deep networks
  - Connections to neuroscience
- Complex systems
  - How can we describe (aspects of) complex systems?
  - Emergent structure / order
  - Relation to natural science




STRUCTURE


COMPLEXITY

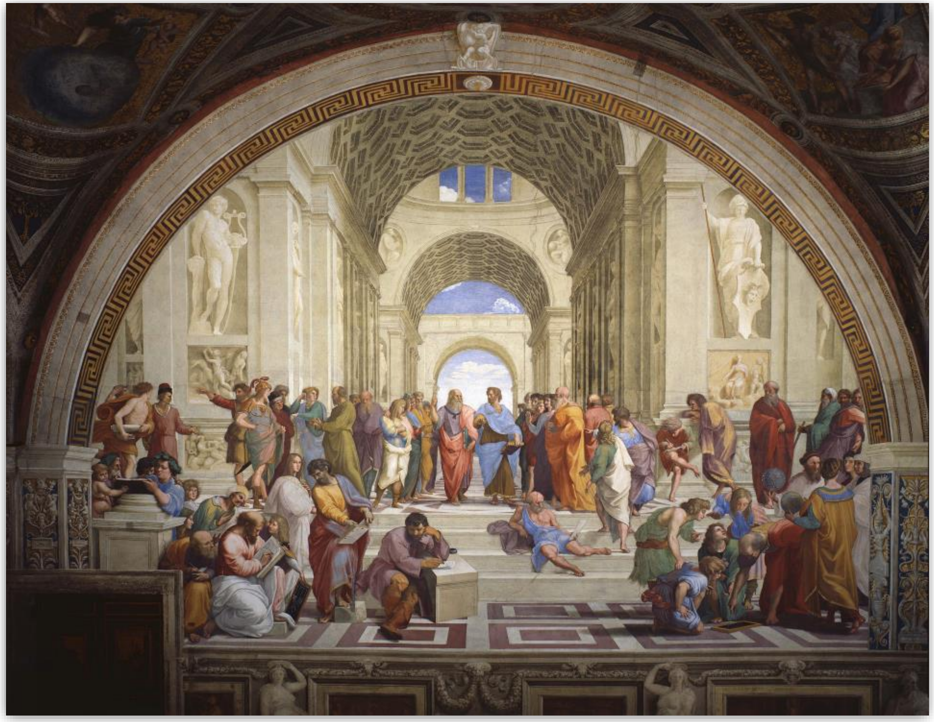
Knowledge?

How do we know things?

Epistemology (“Erkenntnistheorie”)

- How do we get to know things
- And be *reasonable* to be *reasonably sure* about them

How do we know things?

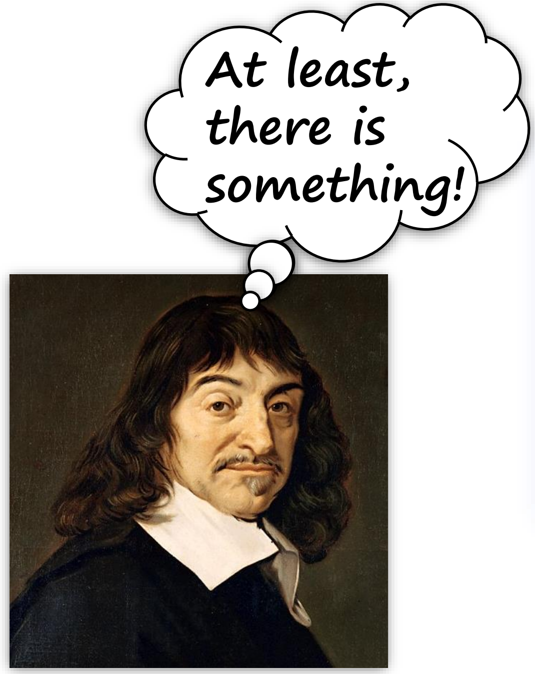


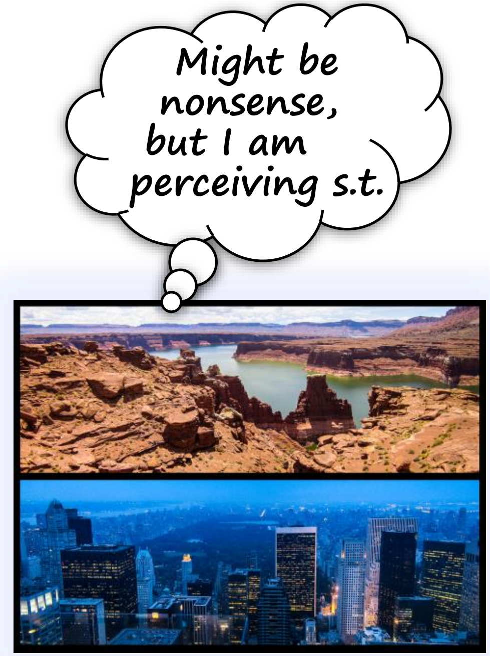
Socrates

- Skepticism: “I know that I know nothing”
- **Basis of all science** (but *clearly* insufficient on its own…)

[Raffael. The Yorck Project (2002), public domain. *https://commons.wikimedia.org/w/index.php?curid=157721]*

How do we know things?





Descartes

- “Cogito ergo sum” – I think, therefore I am
- Consciousness: Important, but not our topic

[https://commons.wikimedia.org/wiki/File:Frans_Hals_-_Portret_van_Ren%C3%A9_Descartes.jpg]

The Science of Knowing

To make progress

- Need "philosophically" strong assumptions
- Not "strong" in an everyday-sense

Historic: Enlightment

- Let's be reasonable
  - But what is reasonable?

[Sébastien Leclerc I, Louis XIV Visiting the Royal Academy of Sciences. https://commons.wikimedia.org/w/index.php?curid=30184020]

The Scientific Method

Assumptions

- Math & Logic
  - Occasionally non-trivial
  - See e.g. debate on “axiom of choice”
- Symmetry
  - Repeatability of experiments
  - Spatio-temporal persistence of knowledge
- Simplicity
  - How the world works can be condensed to a few “simple” rules
  - “Reductionism”

The Scientific Method

Gaining objective knowledge

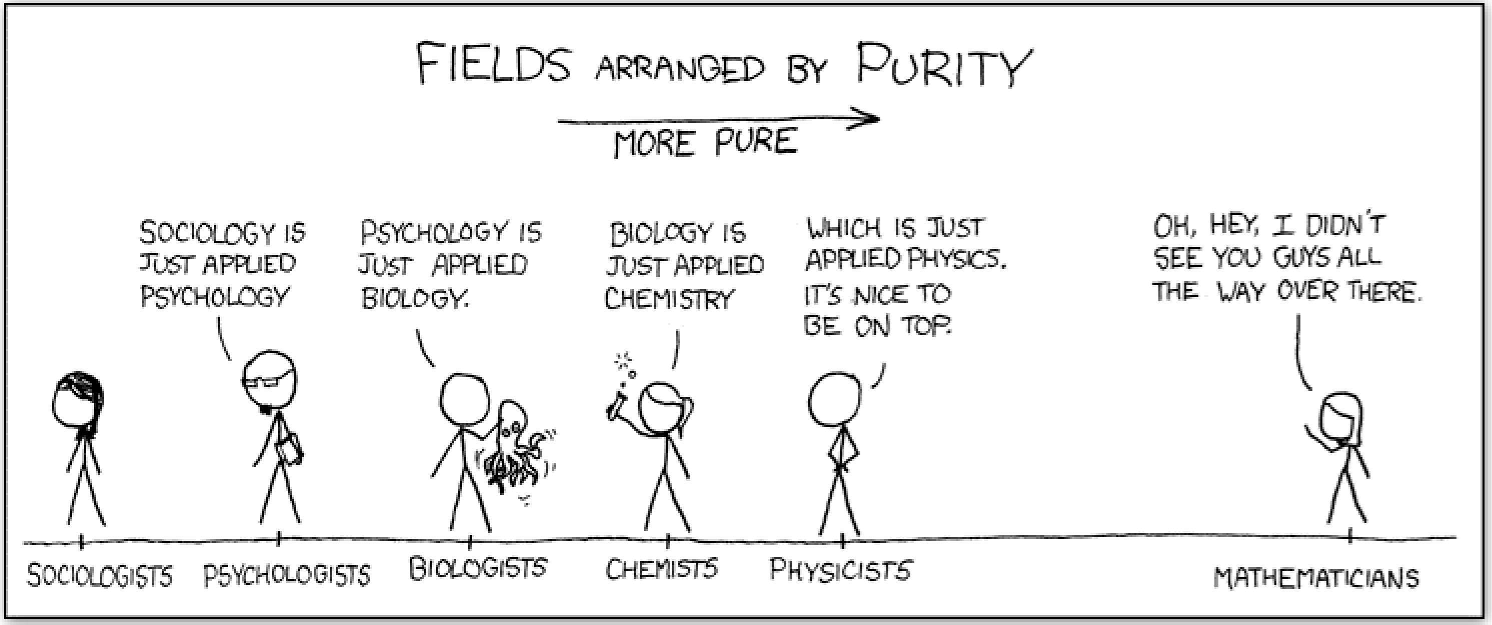
- Various formulations
  - Discussing my personal take here
- Skepticism as default

Two main techniques

- Logical reasoning (deduction)
- Empirical observation (induction)

(At least)

Two Schools of Thought



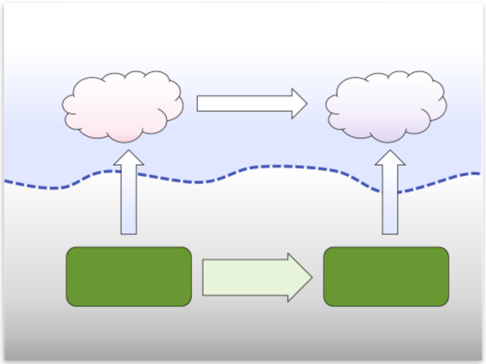
empirical

purely deductive

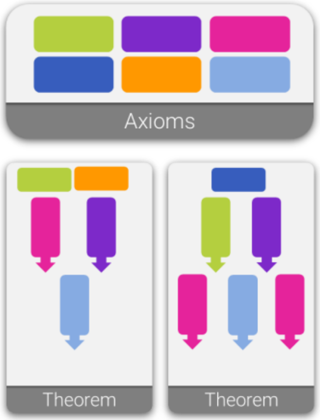
Comic: [Randall Munroe, https://xkcd.com/435/]

(At least)

Two Schools of Thought



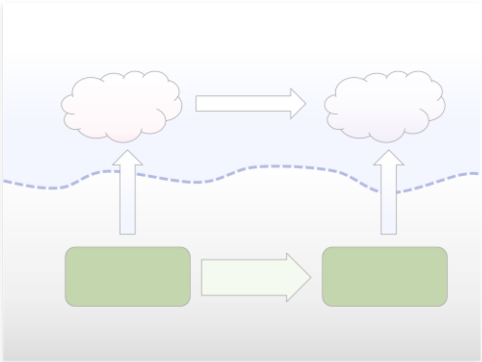
Induction



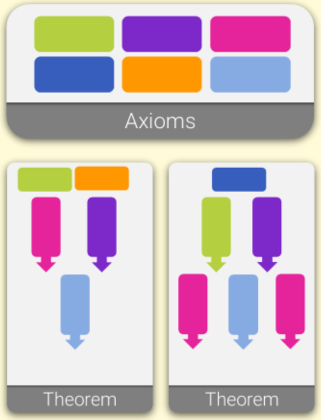
Deduction

(At least)

Two Schools of Thought



Induction



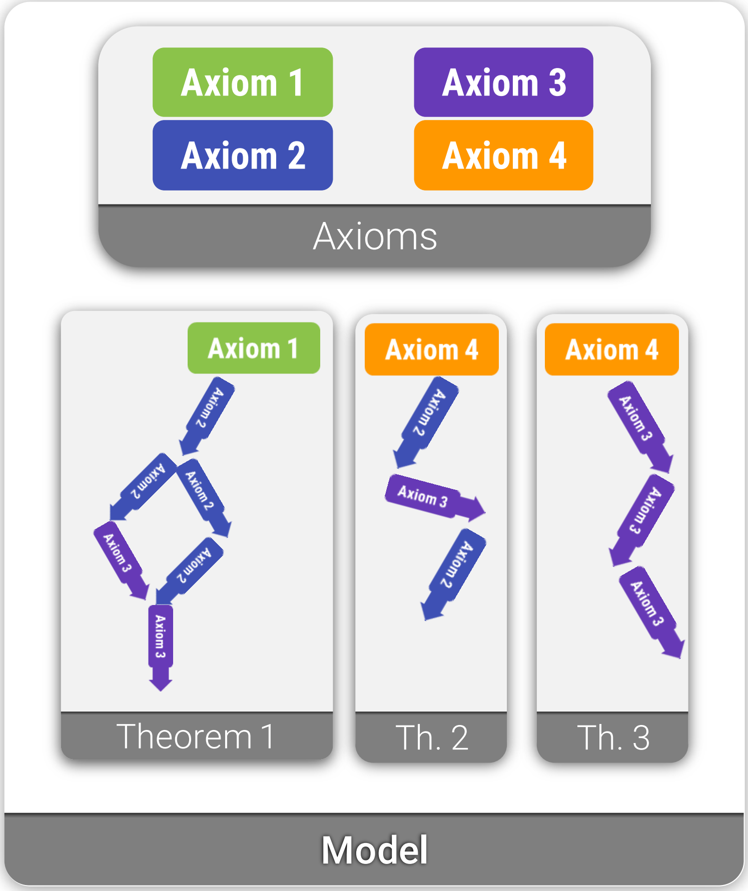
Deduction

Deductive Reasoning

Deductive Reasoning

- Start from assumptions
  - “Axioms”
- Derive consequences

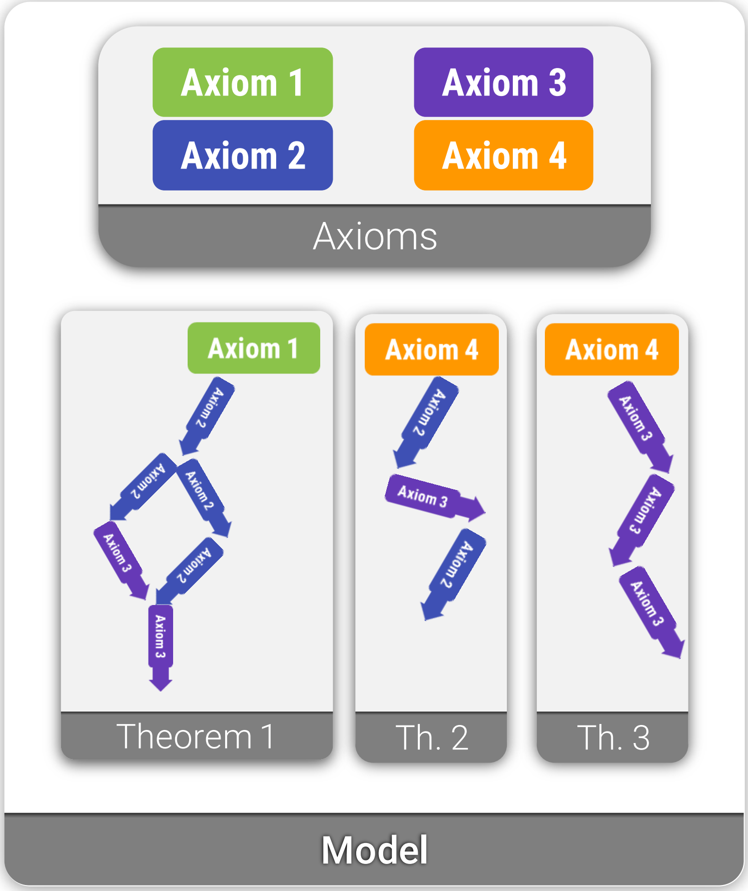
≈ “Mathematics”



Deductive Reasoning

Structure

- Sequence of invocations of assumed facts yields new facts
- Can be complex
  - Variables (and sets)
  - Higher-order logic



Example

Peano Axioms for $\mathbb{N}$ *(excerpt)*

- `0` is a natural number
- For each $n \in \mathbb{N}$ there is a successor $S(n)$
- If $n = m$ then $S(n) = S(m)$
- …

Complexity: Use of variables

- Making statements over variables from larger sets
  - Expressive power depends on types of sets permitted
  - E.g. sets of sets vs. single elements

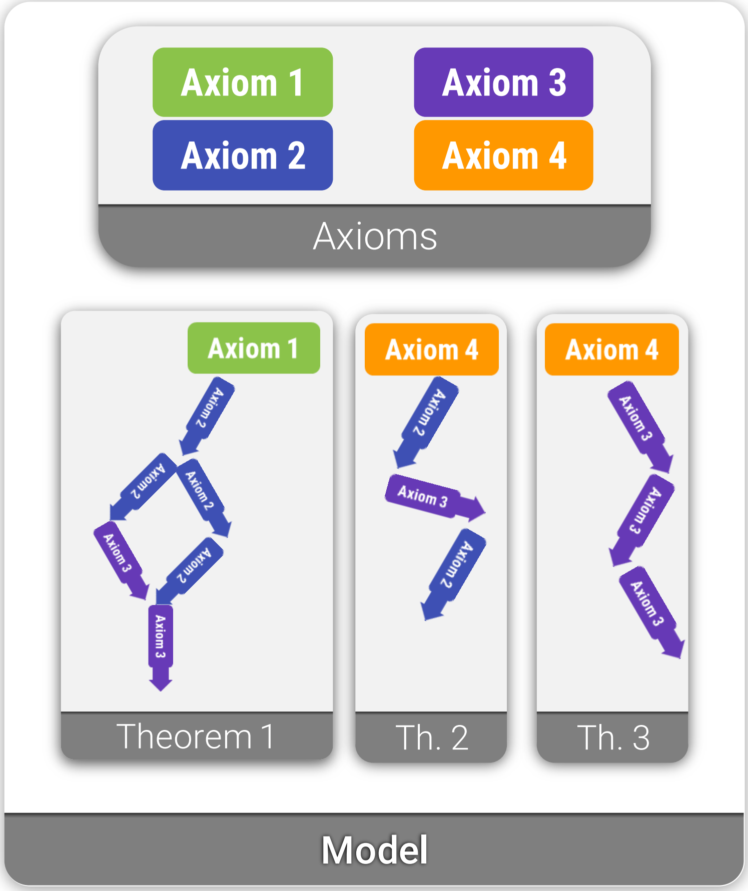
Deductive Reasoning

Computational Structure

- Axioms and proofs can be encoded as bit-strings
- Countably many proofs

Automatic proving

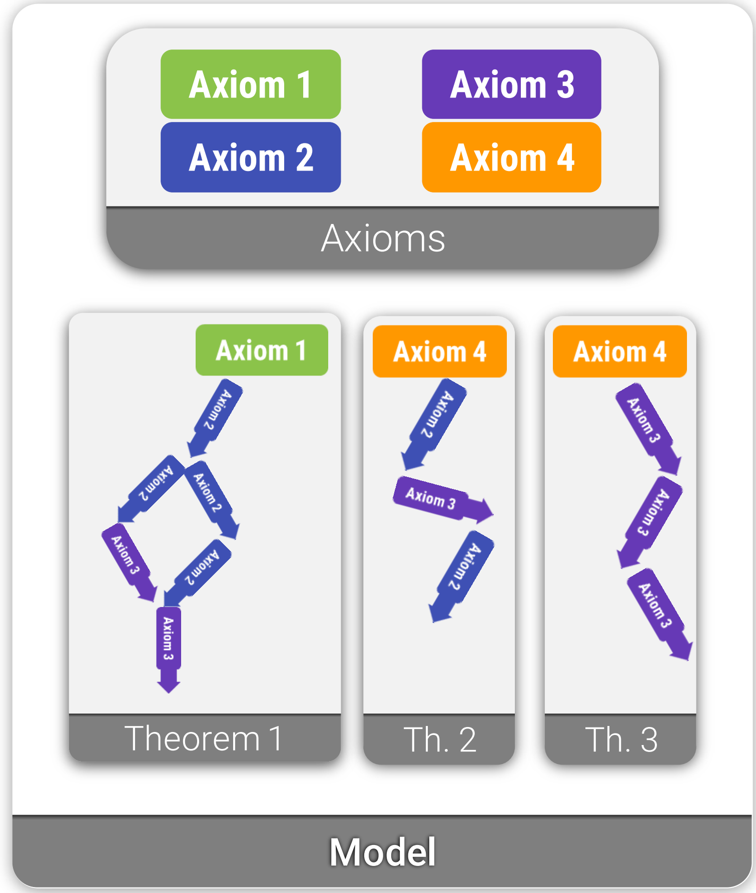
- Algorithmic deduction (part of AI, but not our topic)
- We can search for proofs
  - Undecidable: no exclusion of existence in finite time
  - Means: very, very expensive search in practice



It gets worse…

Gödel's incompleteness

- Axioms strong enough to describe ℕ
  - „All facts“ are not recursively enumerable, but proofs are



Consequence

- There are “true” facts without a proof
  - For classic binary logic
- **Emergent complexity:** You cannot understand an axiom system from within itself

Speaking of Static Logic…

Emergent complexity in algorithms

- "Dynamically" executed algorithms face the same problem

```
// good night & good luck.
int f(int n) {
    if (n <= 1) return 0;
    if (n % 2 == 0) {
        return f(n / 2);
    } else {
        return f(3*n + 1);
    }
}
```

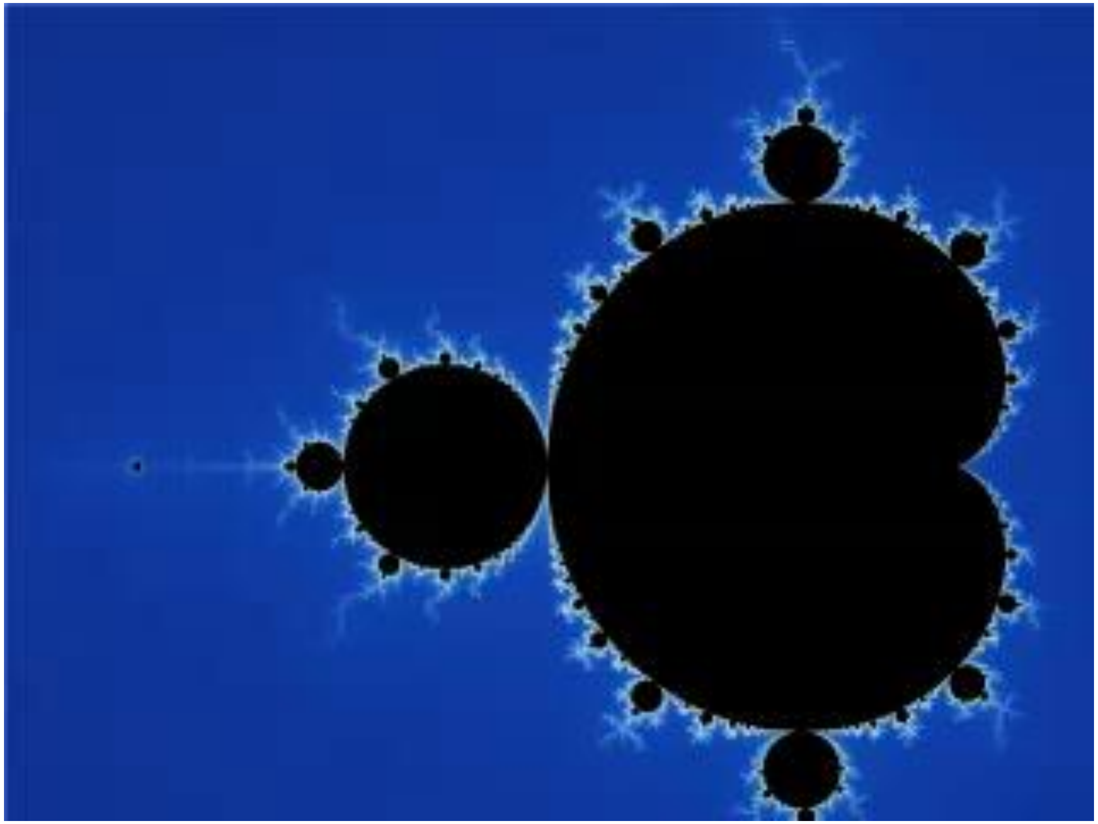
Unpredictable behavior

- Turing-capable program
  - E.g. arithmetics, assignment, repetition, condition
  - Discrete, finite (but unbound) sequence of statements
- No finitely-sized algorithms can decide non-trivial properties of the algorithms behavior

Emergent Complexity

Emergent Complexity

$\text{im}(c) \rightarrow$



$\text{re}(c) \rightarrow$

Mandelbrot Set

Iteration

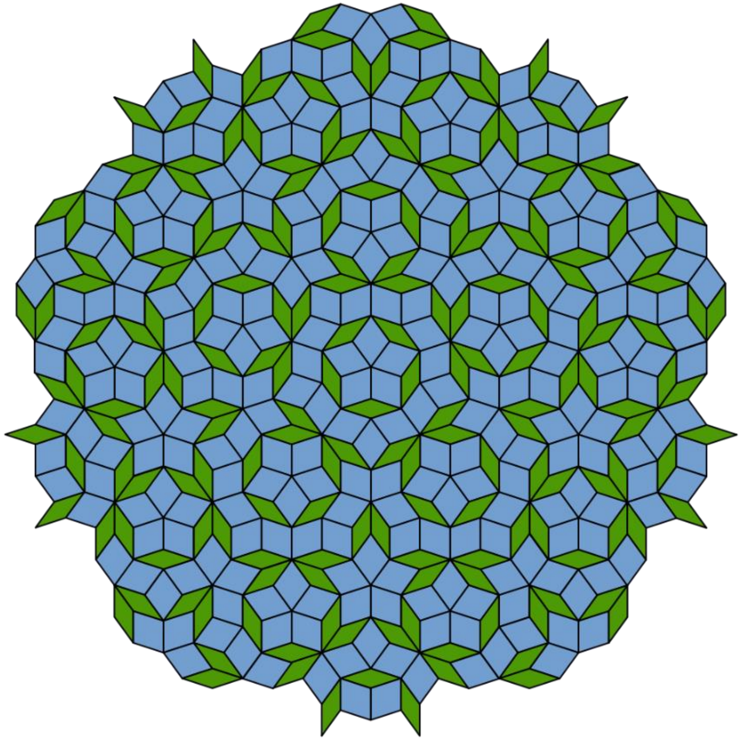
$z \rightarrow z^2 + c$

$z, c \in \mathbb{C}$

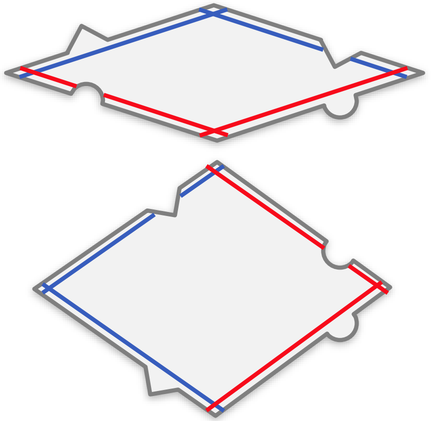
color = number of iterations until value > 2

[Animation: Wikipedia contributor „Simpsons contributor"]

Emergent Complexity



Penrose Tilings



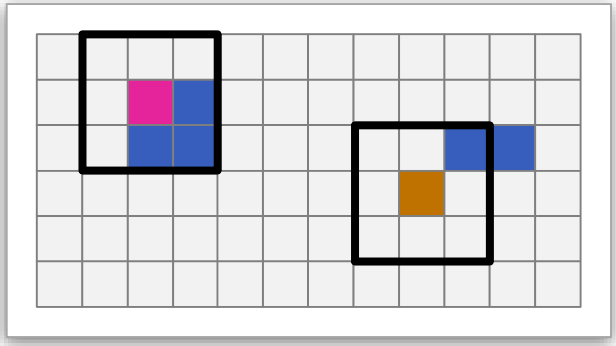
Principle

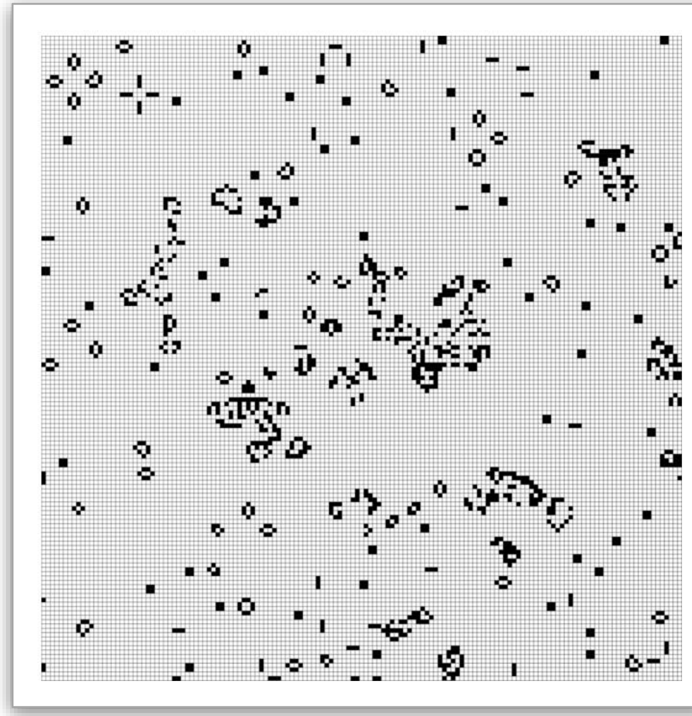
- Put tiles together
- Can fill whole 2D plane
- Globally aperiodic
  (no repetitions)

Tiling is Turing-capable → Programming by designing tiles
(quadratic “Wang-Domino” bricks sufficient)

[Left tiling: WP contrib. Inductiveload, https://en.wikipedia.org/wiki/Penrose_tiling]

Emergent Complexity





Conway‘s „Game of Life“

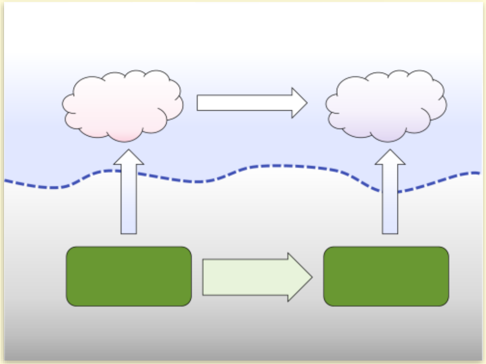
- Binary lableing for „cells“: dead or alive
- “Living” cells with <2 or >3 neighbors “die”
- “Living” cells with 2…3 neighbors remain “alive”
- “Dead” cells with exactly 3 neighbors come back to live

Turing-capable machine

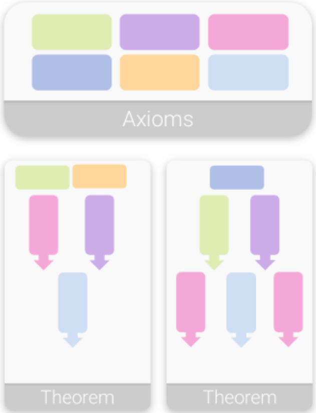
[right image: Jan Disselhof]

(At least)

Two Schools of Thought



Induction

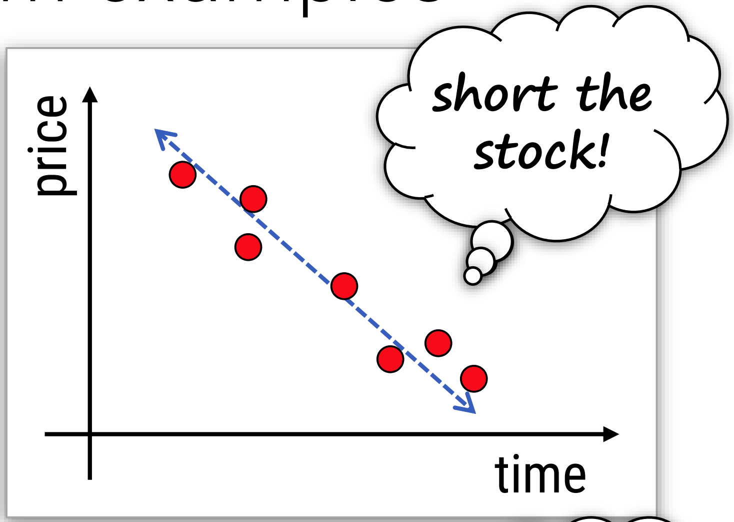


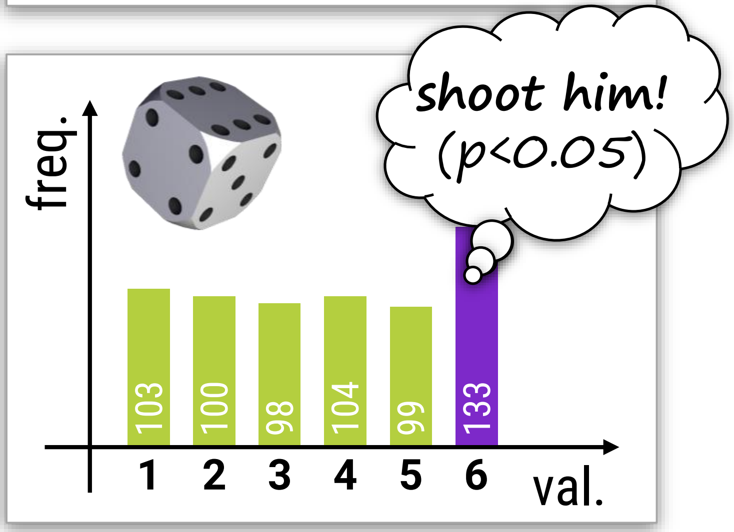
Deduction

Induction*)

Induction: Generalization from examples

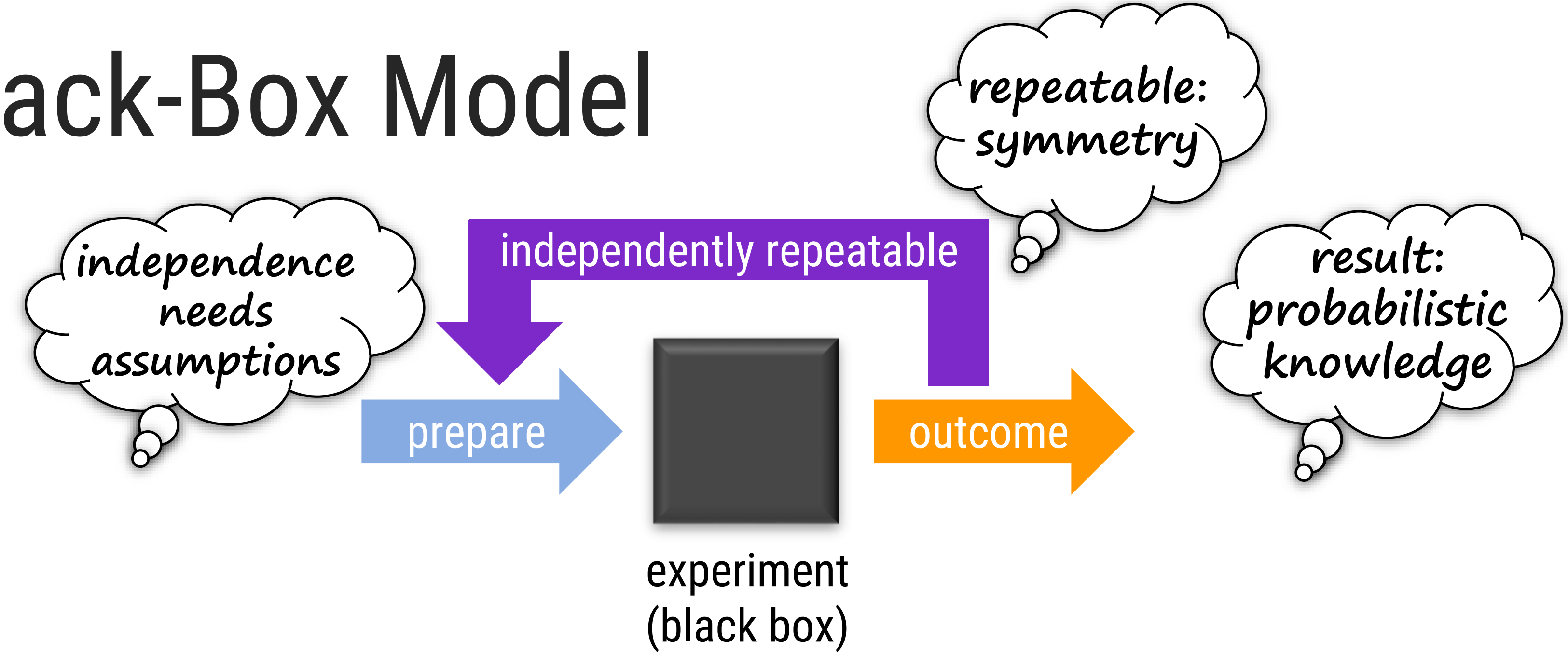
- Observe examples
- Try to derive model
- Observe more (independent) examples
- Verify or falsify model





*) advisory: do not mix up with “proof by induction”

Black-Box Model



Repeatable / reproducible experiment

- Define/find/observe experiment that is repeatable
  - Same behavior each time (rest "randomness")
  - Independent (no influence, *also* not in randomness)
- Make multiple observations
- Gain information on how likely outcomes are

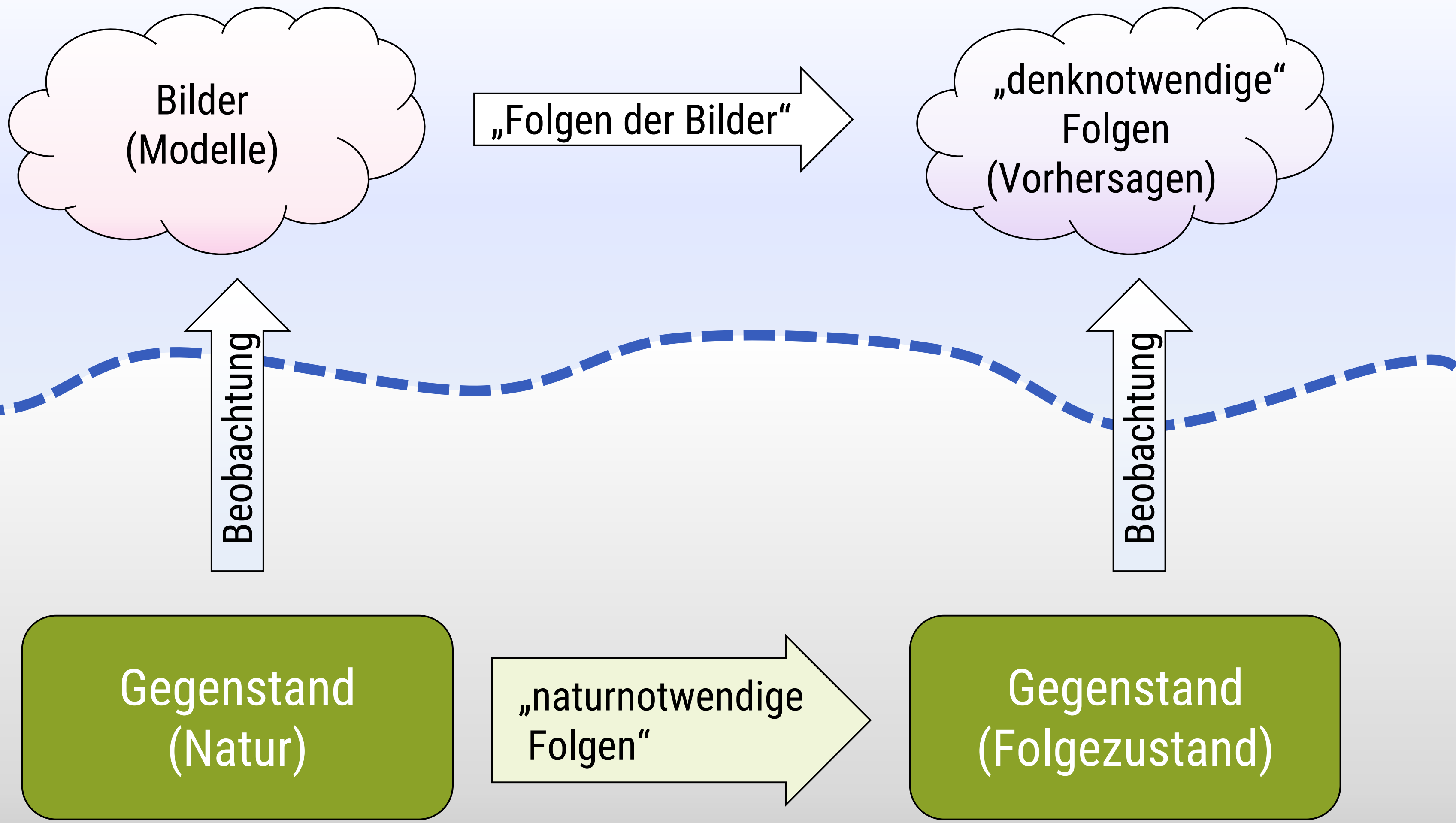
Induction Recipe

How to learn knowledge inductively

- Set up model
  - Might contains unknown parameters
- Find repeatable experiment (“training data”)
- Measure n outcomes
- Determine
  - If the model is able to explain the experiment
  - Which parameters are likely

realm of ideas
Models,
Imagination

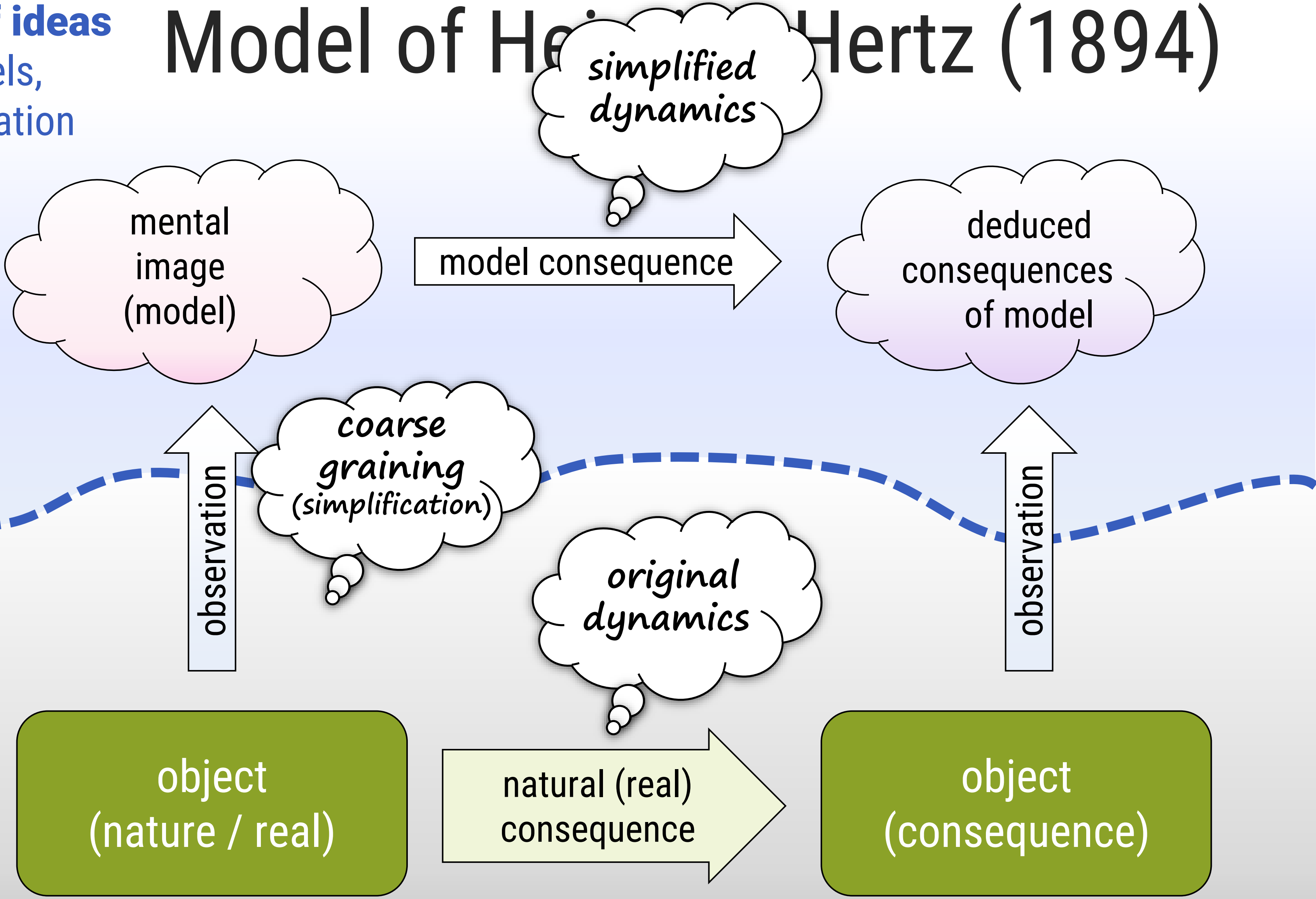
Model of Heinrich Hertz (1894)



real world (objective & unknown)

realm of ideas
Models,
Imagination

Model of He… Hertz (1894)

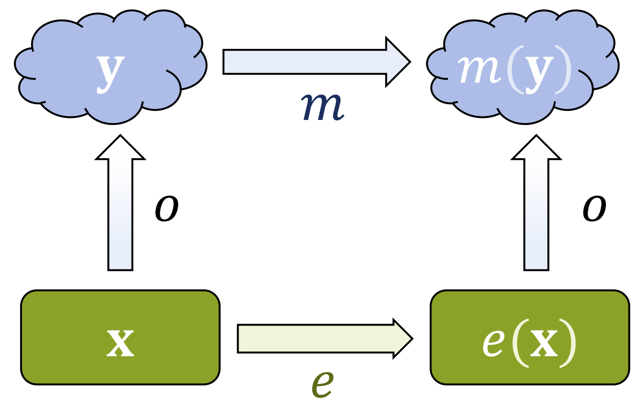


real world (objective & unknown)

Formalization

Inductive / predictive reasoning

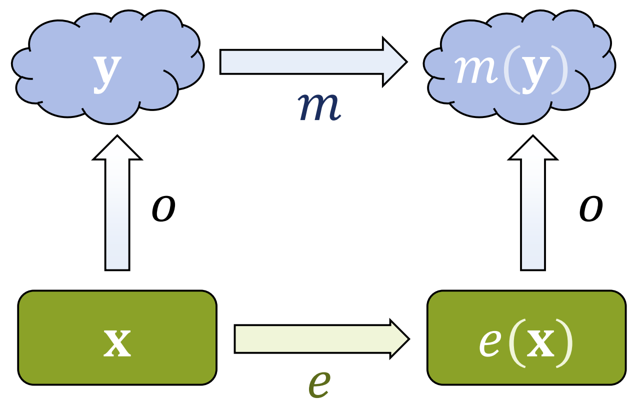
- "State of nature": $\mathbf{x} \in \Omega$
- "State of model": $\mathbf{y} \in \mathrm{M}$
- Observation: $o\colon \Omega \to \mathrm{M}$
- Experiment: $e\colon \Omega \to \Omega$
- Model prediction: $m\colon \mathrm{M} \to \mathrm{M}$



Commuting diagram / homomorphism

- Chose m such that $\forall \mathbf{x} \in \Omega\colon m \circ o = o \circ e$

Formalization



Learning

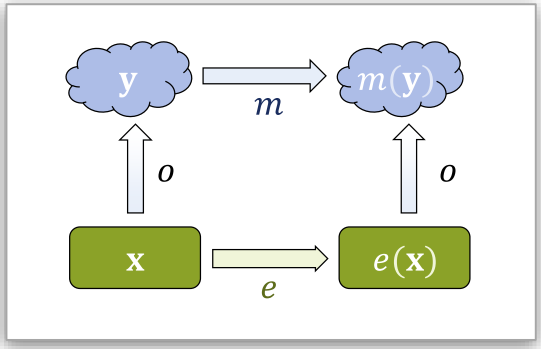
- Chose m such $\forall \mathbf{x} \in \Omega: m \circ o \approx o \circ e$ unkown

Inference (using knowledge)

- Find $y' \in M$ such $y' = m(o(\mathbf{x})) = m(y)$ unkown

What Could Possibly Go Wrong?

Is this model sufficient?

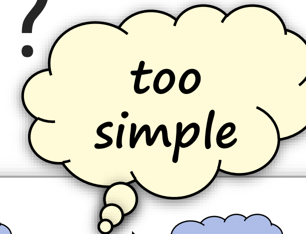


- Assume perfectly predictive model
- Is this model "correct"?

$2\frac{1}{2}$ Problems

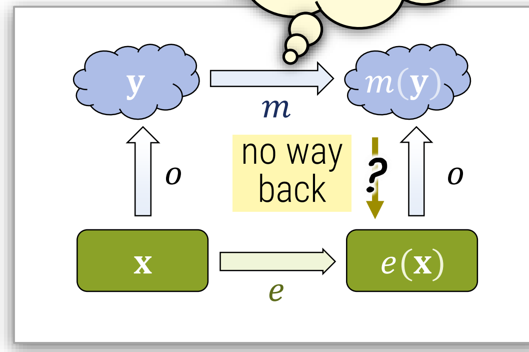
- The model might be too complex — bad
- The model might be (overly) simplified — can be ok
- Information is probabilistic — unavoidable (need math)

What Could Possibly Go Wrong?



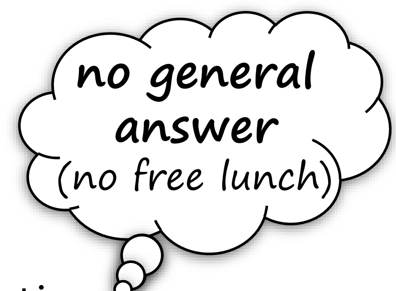
(1) Too simple

- Observation / model might loose information
- Model $\mathbf{y}$ does not describe all of $\mathbf{x}$



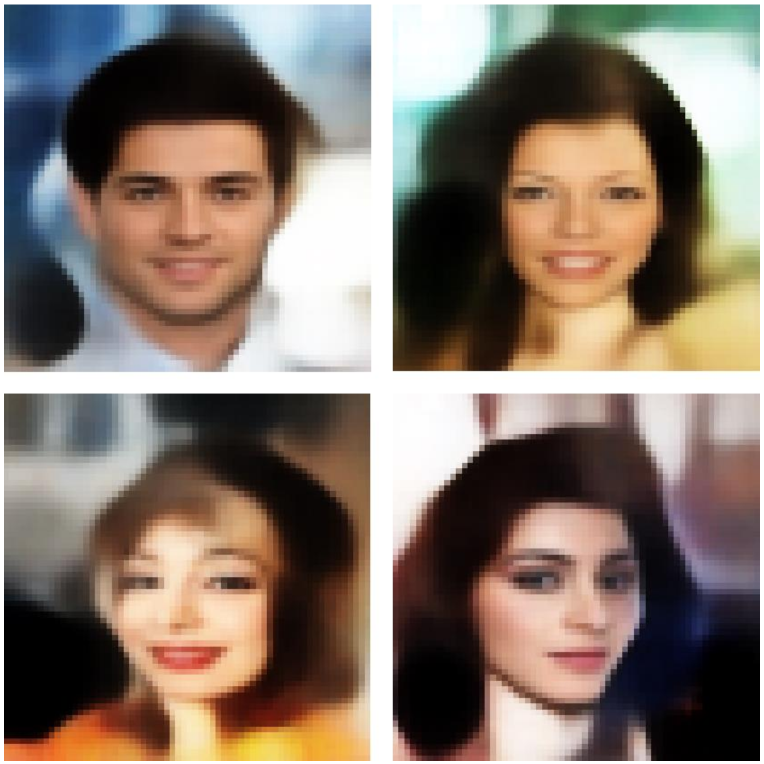
"In principle ok"

- No model is comprehensive
  - Abstraction needed
  - Need to keep the "right" / "relevant" information

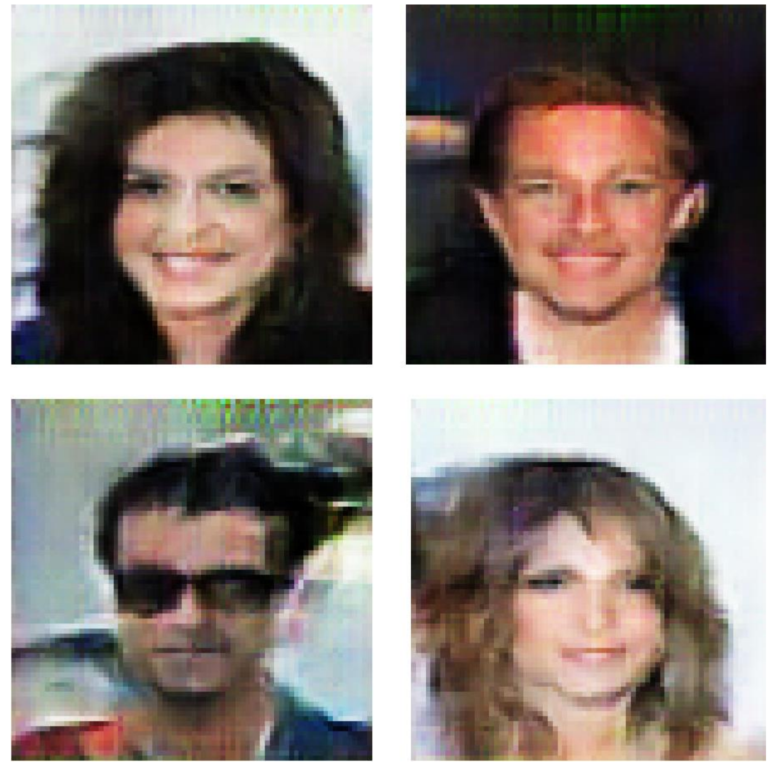


- Model design problem
  - Example: image reconstruction vs. pattern recognition

Model might leave out details…



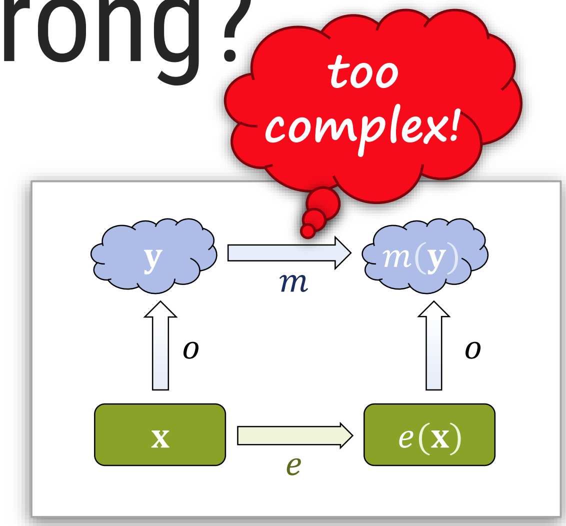
Autoencoder
(PCA in latent space)



WGAN-GP
(generative adversarial network)

[results courtesy of D. Schwarz, D. Klaus, A. Rübe]

What Could Possibly Go Wrong?



(2) Too complex

- Observation (model) has too much information
  - Model might *add* information
  - Model might not *remove enough* information

Usually: Very bad

- Additional information is nonsense
  - “Made-up stuff”, “Fairytale”
- Predictive power might be compromised
  - If not careful, we might not recognize it

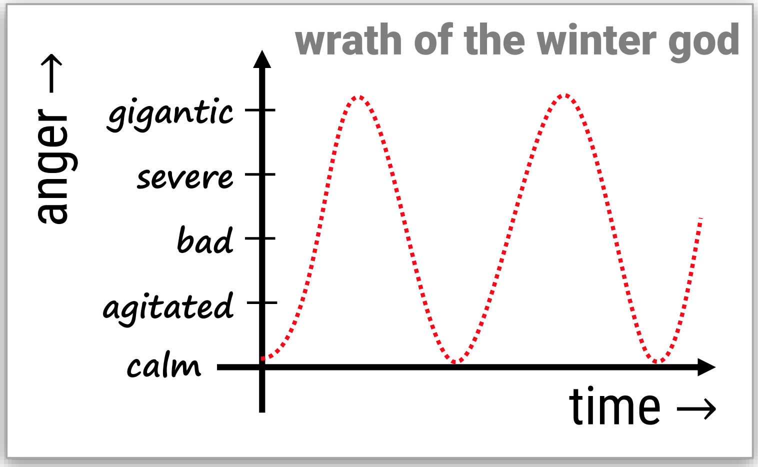
Example: Mythology of Seasons

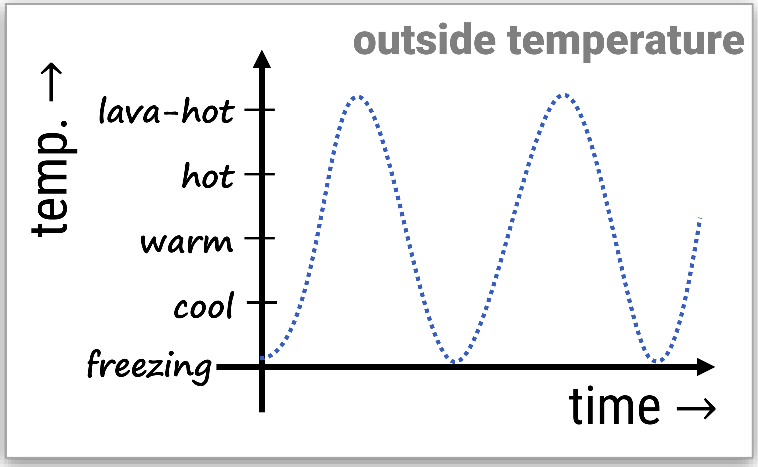
Example of "too complex" [see Ref. below]

- The sun is a goddess. Shines warm.
  - (Note: latent sexism in mythology)
- There is a winter god. He is evil and moody.
  - Why? This is how bad guys are in mythology!
- When the winter good gets in a bad mood, he chases the sun in senseless wrath.
  - The sun has to hide.
  - It happens periodically.

Adapted from David Deutsch: "A new way to explain explanation" (Ted Talk 2009)
[https://www.ted.com/talks/david_deutsch_a_new_way_to_explain_explanation]

Model too Complex



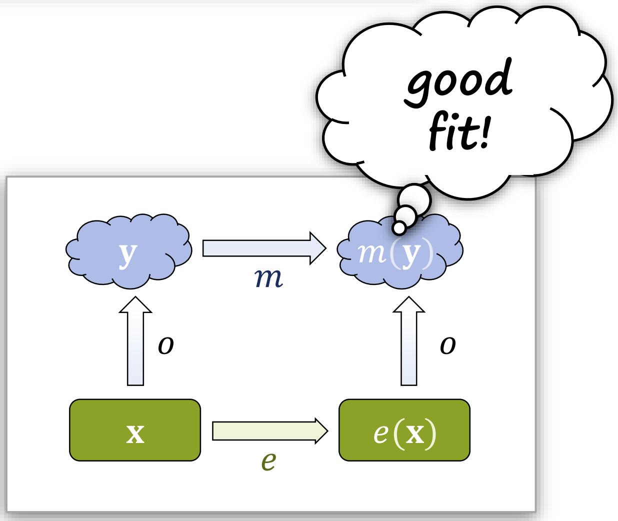


Good model?

- Explains seasons very accurately

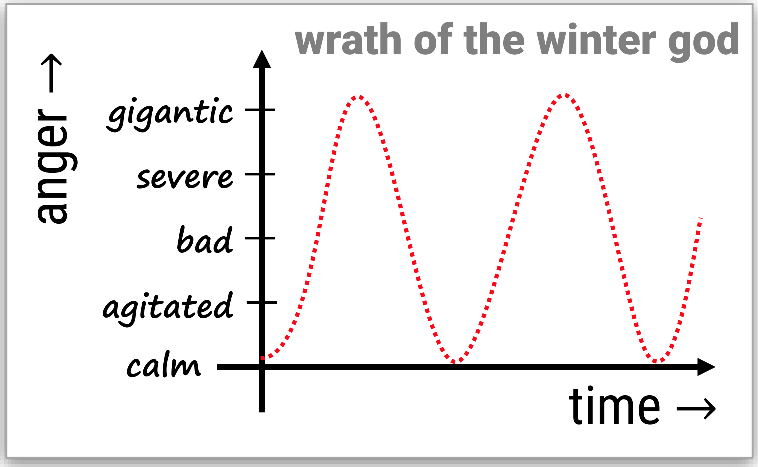
$$\|m(o(x)) - o(e(x))\| \rightarrow small$$

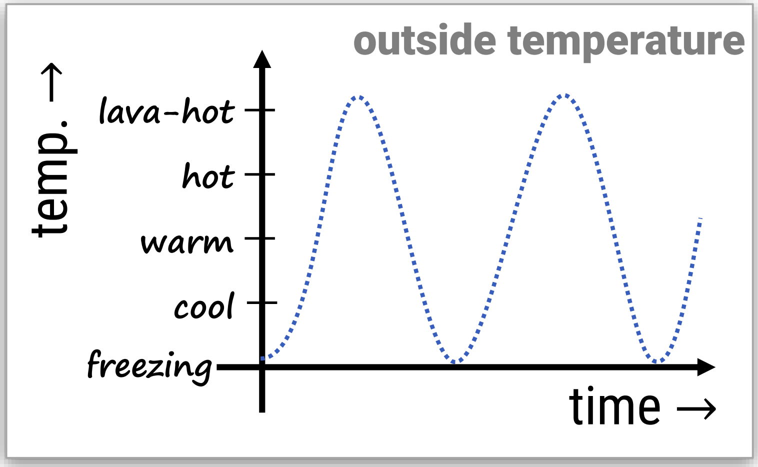
- All predictions match very well



Adapted from David Deutsch: "A new way to explain explanation" (Ted Talk 2009)
[https://www.ted.com/talks/david_deutsch_a_new_way_to_explain_explanation]

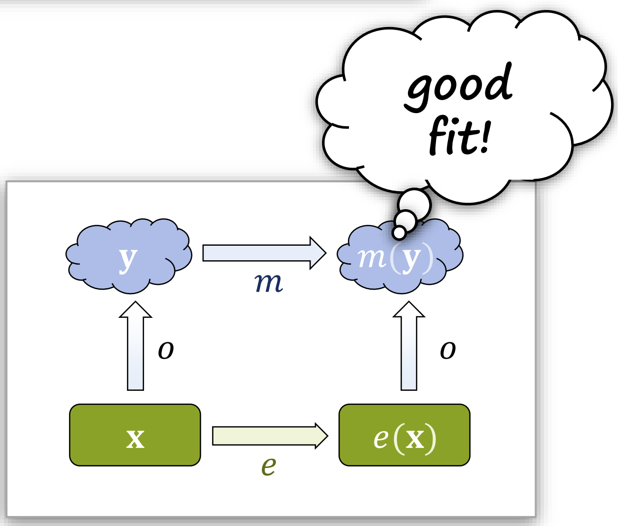
Model too Complex





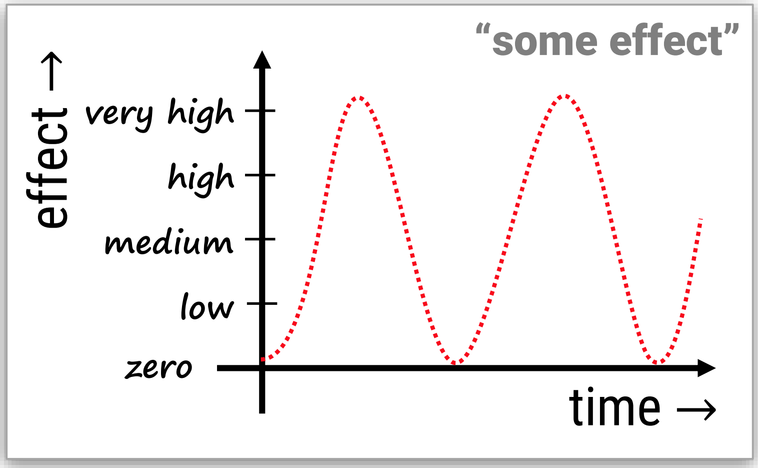
Good model?

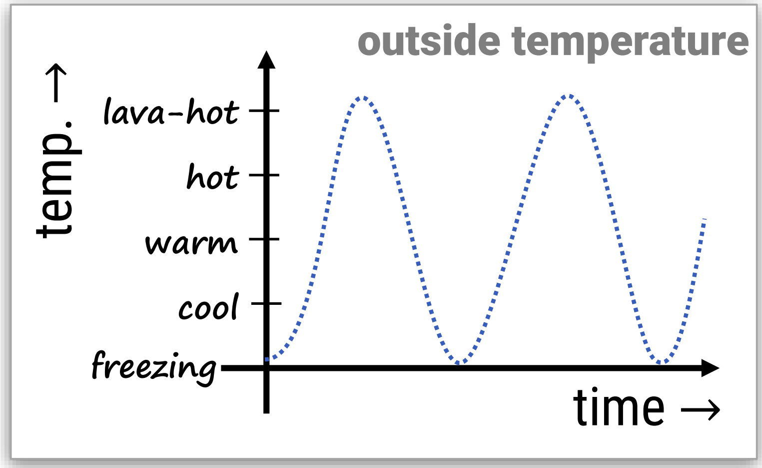
- Unverifiable model aspects
  - „Winter god" → "guys next town release poisonous gas into the air"
- Unreliable information



Adapted from David Deutsch: "A new way to explain explanation" (Ted Talk 2009) [https://www.ted.com/talks/david_deutsch_a_new_way_to_explain_explanation]

Model too Complex



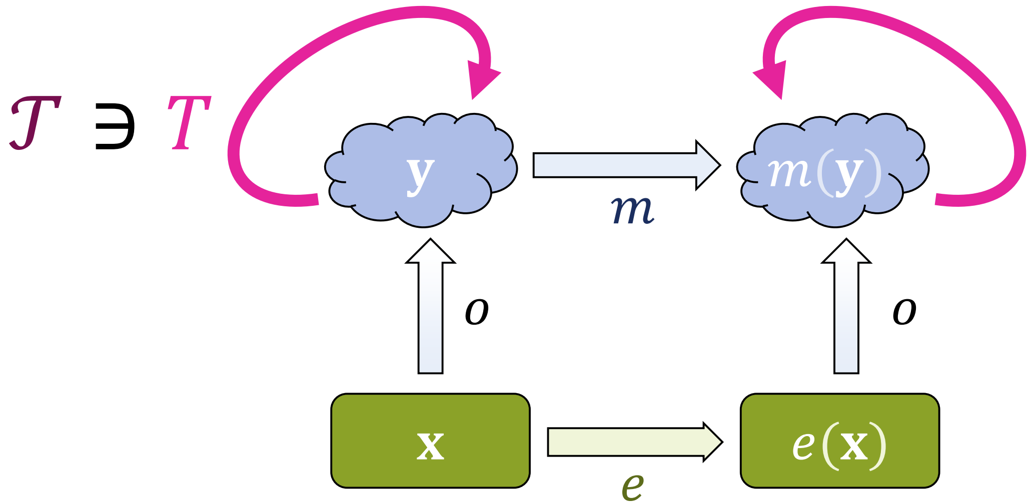


Rule: Falsifiable models

- Remove all information that is independent of *experiment*-*observation*-cycle
- Anything that can be changed without changing the outcome is "no information"

Adapted from David Deutsch: "A new way to explain explanation" (Ted Talk 2009) [https://www.ted.com/talks/david_deutsch_a_new_way_to_explain_explanation]

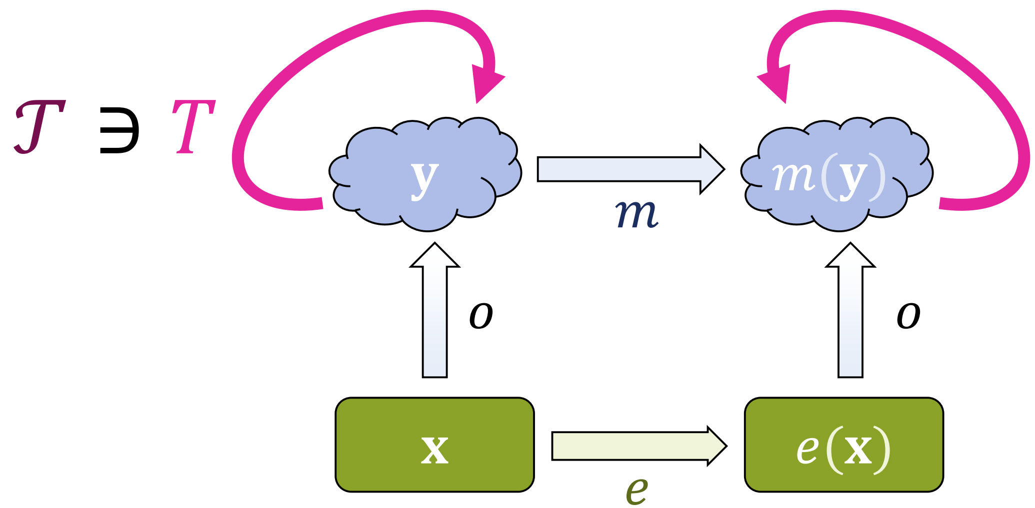
Model too Complex



What we do not know

- Let T be a change to the model
- Anything change that does not change the observations yields an equally predictive model
- All the information subject to change are "not established" (unknown)

Symmetry in overly complex models



Formally: Symmetry

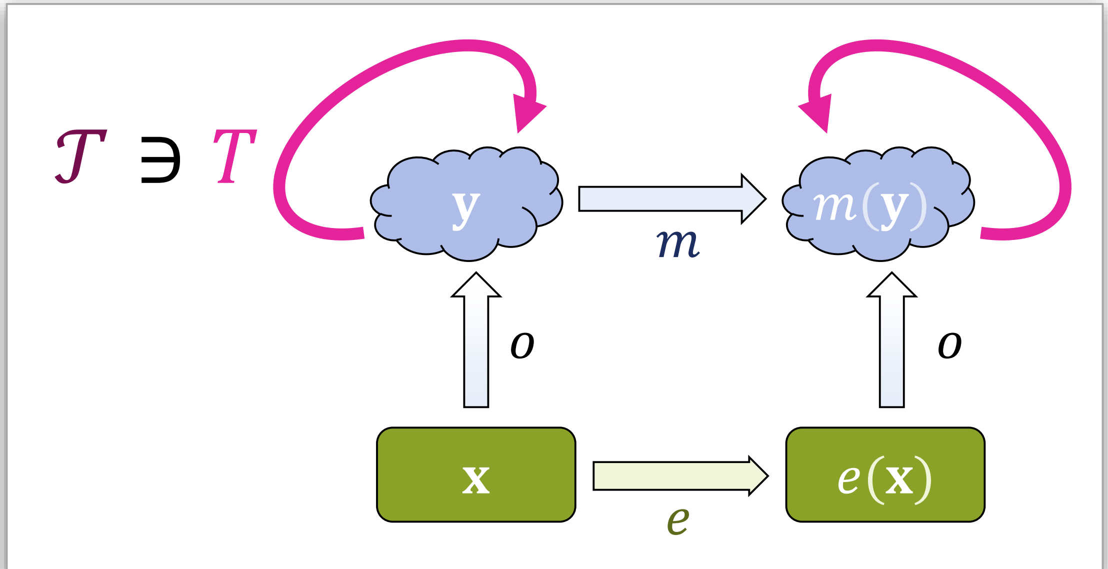
- Transformation $T \in \mathcal{T}$ leave experiments unchanged

$$T: M \to M$$

$$\mathcal{T} = \{T: M \to M \mid \forall \mathbf{x} \in \Omega: m \circ T \circ o \approx T \circ o \circ e\}$$

- $\mathcal{T}$: *symmetry group* of the *model* under *observations*
- We do not know $\mathbf{y}$, only $\mathbf{y} \bmod \mathcal{T}$

Symmetry in overly complex models

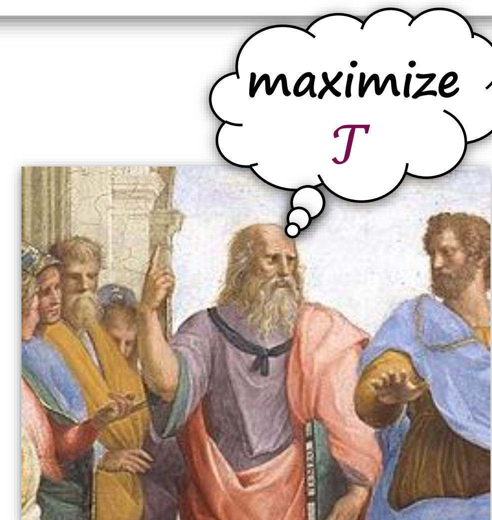


$\mathcal{T}$ = set of all permutations of M:

$$\forall y_1, y_2 \in M:$$
$$y_1 \equiv y_2 \bmod \mathcal{T}$$

Socrates: Do not believe anything

- My thinking might be delusional
- Observations might be hallucinations
- No knowledge: All models equal



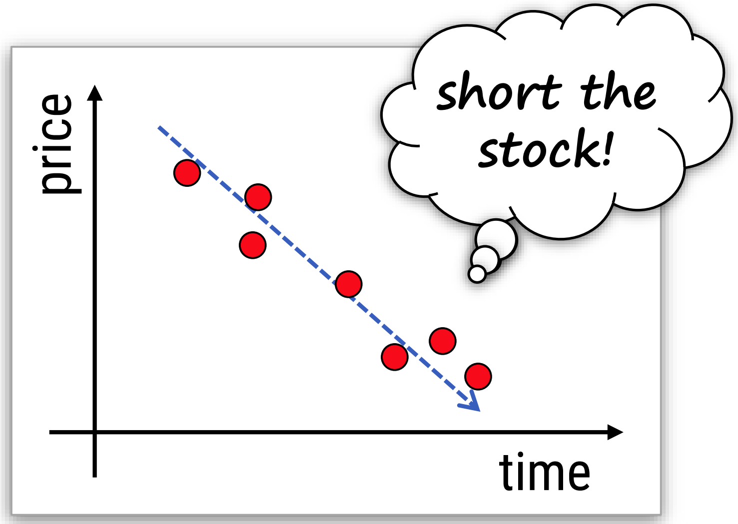
Unreliable Models

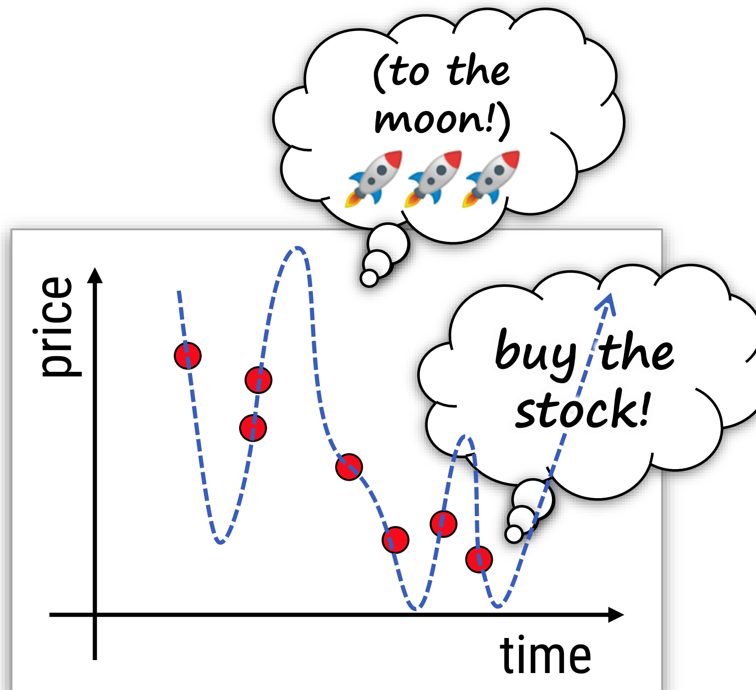
Another example

- Betting on stock prices
- Polynomial fitting
- Seven observations

Degree k polynomial

- $k = 6$ fits any data
  - Unique model
  - But no predictive power
- $k = 5,4,3$ ...? fits any data
  - More or less reliable





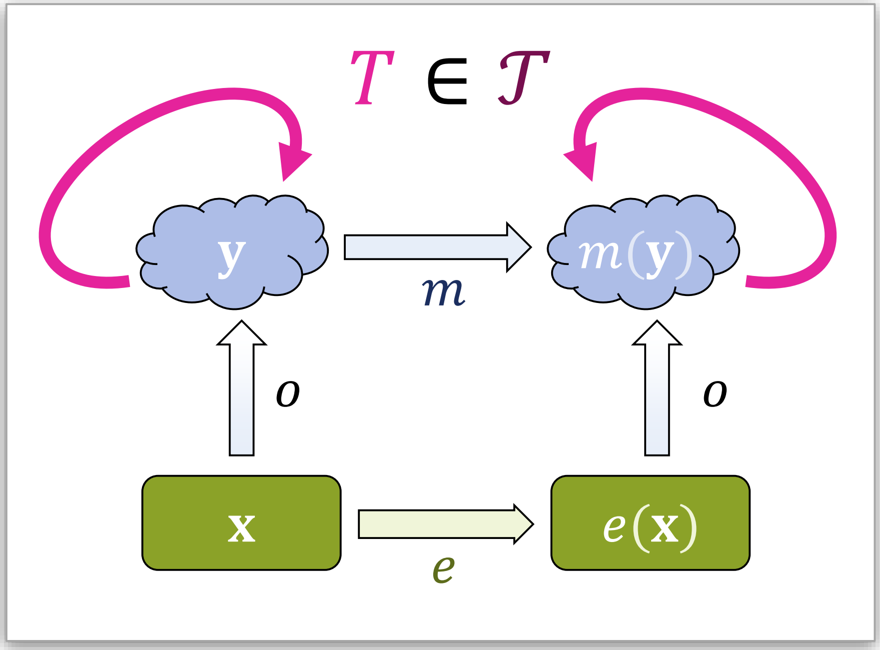
Unreliable Models

Another example

- Betting on stock prices
- Polynomial fitting
- Seven observations

Degree k polynomial

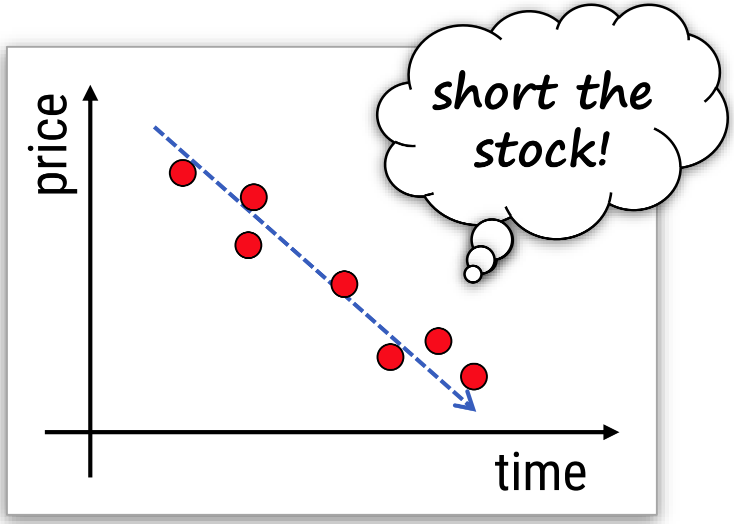
- $k = 6$ fits any data
  - Unique model
  - But no predictive power
- $k = 5,4,3$ ...? fits any data
  - More or less reliable



Unreliable Models

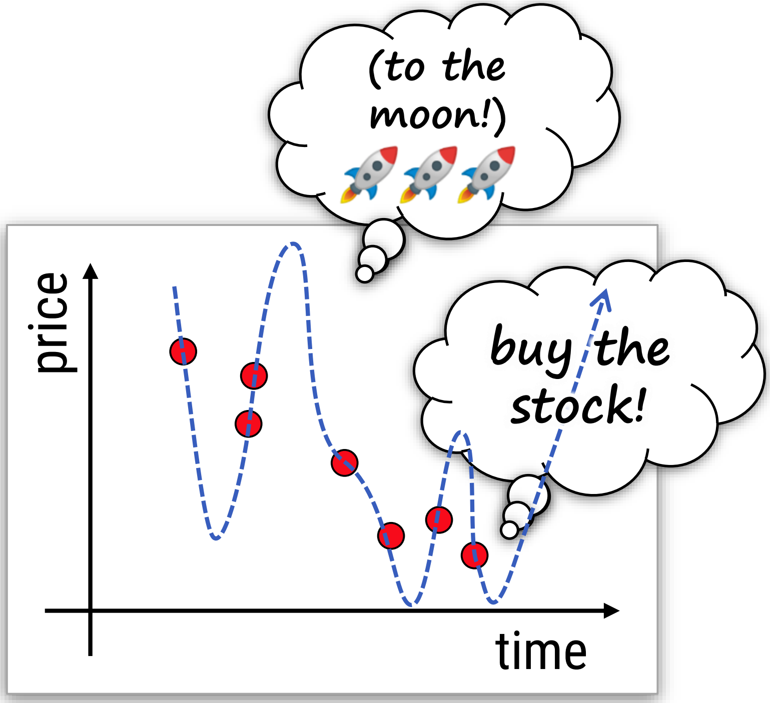
Another example

- Betting on stock prices
- Polynomial fitting
- Seven observations



Degree k polynomial

- $k = 6$ fits any data
  - Unique model
  - But no predictive power
- $k = 5,4,3$ ...? fits any data
  - More or less reliable



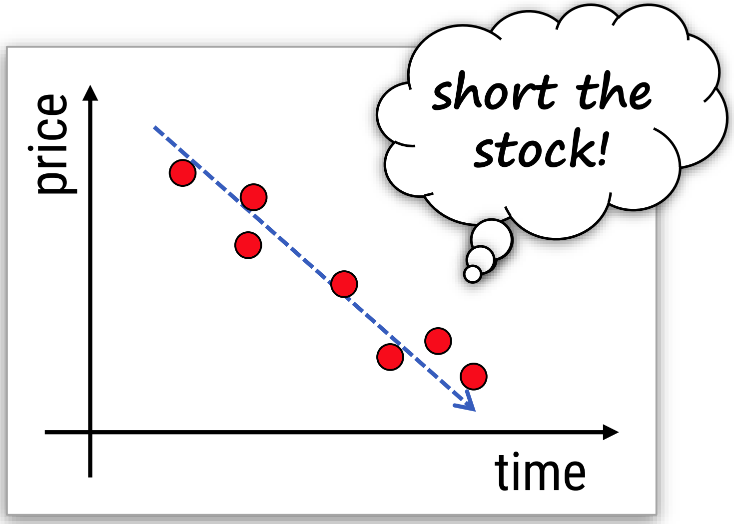
Unreliable Models

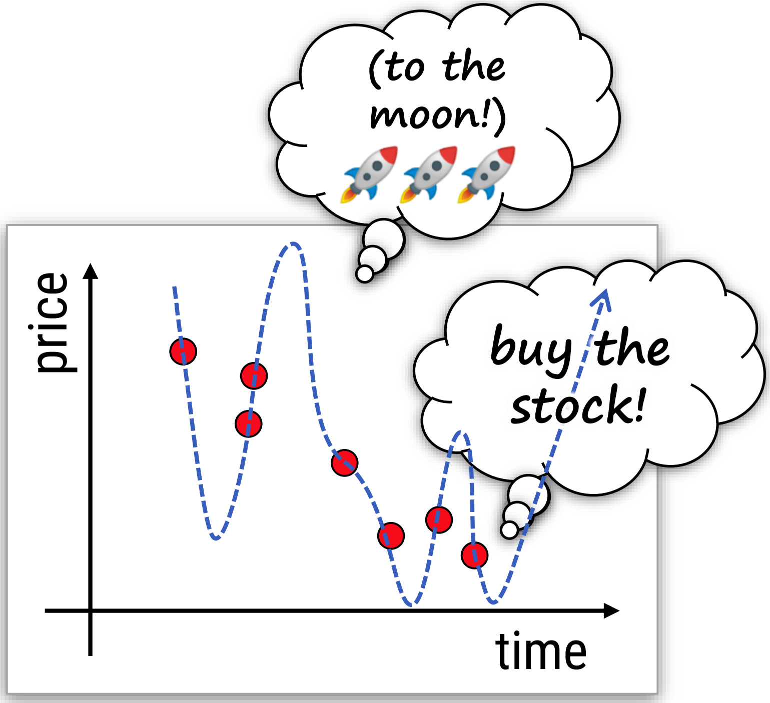
We need to quantify

- How reliable is our model?
- How complex can we make it?

"Occam's razor"

- Do not make overly complex
- We will see a quantitative version soon

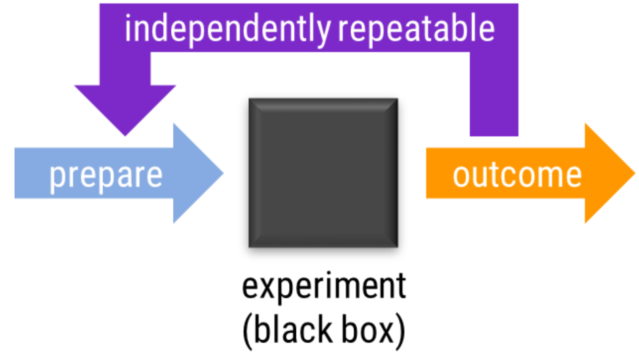




Unreliable Models

Remark

- The basics can still go wrong
- Repeatability / time symmetry

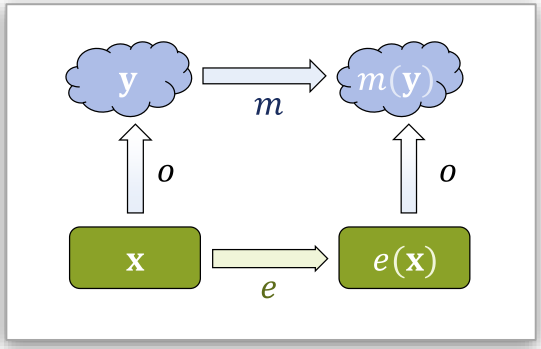


Examples

- Financial crisis 2008 partially attributed to bad risk modeling for credit default correlations
  - “Unlikely that everybody defaults on home loans”
  - Simple model, but fitted to data from growth period
  - “Experiment” not independently repeatable
- Social media 2021 starts discussing stock trades 🚀

Back to our 3 Problems

Is this model sufficient?



$2\frac{1}{2}$ Problems

- The model might be too complex — bad
- The model might be (overly) simplified — can be ok
- Information is probabilistic — unavoidable (need math)

Probabilistic Nature of Induction

(3) Inductive reasoning is always probabilistic

- Same outcome in 1000 experiments?
  - Slim chance of a change the 1001st time
- Cannot make accurate predictions?
  - Random influence on outcome
  - Example:
    - Sometimes the medicine works, sometimes it does not
    - Physiology highly complex
    - Unmodeled effects are “random”

Probabilistic Nature of Induction

(3) Inductive reasoning is always probabilistic

- This is not a fundamental problem
  - Models can be probabilistic
- But we need the right tools to capture uncertainty

→ **Stastistical Data Modeling**

Probability!

Probability

Discrete probability measure

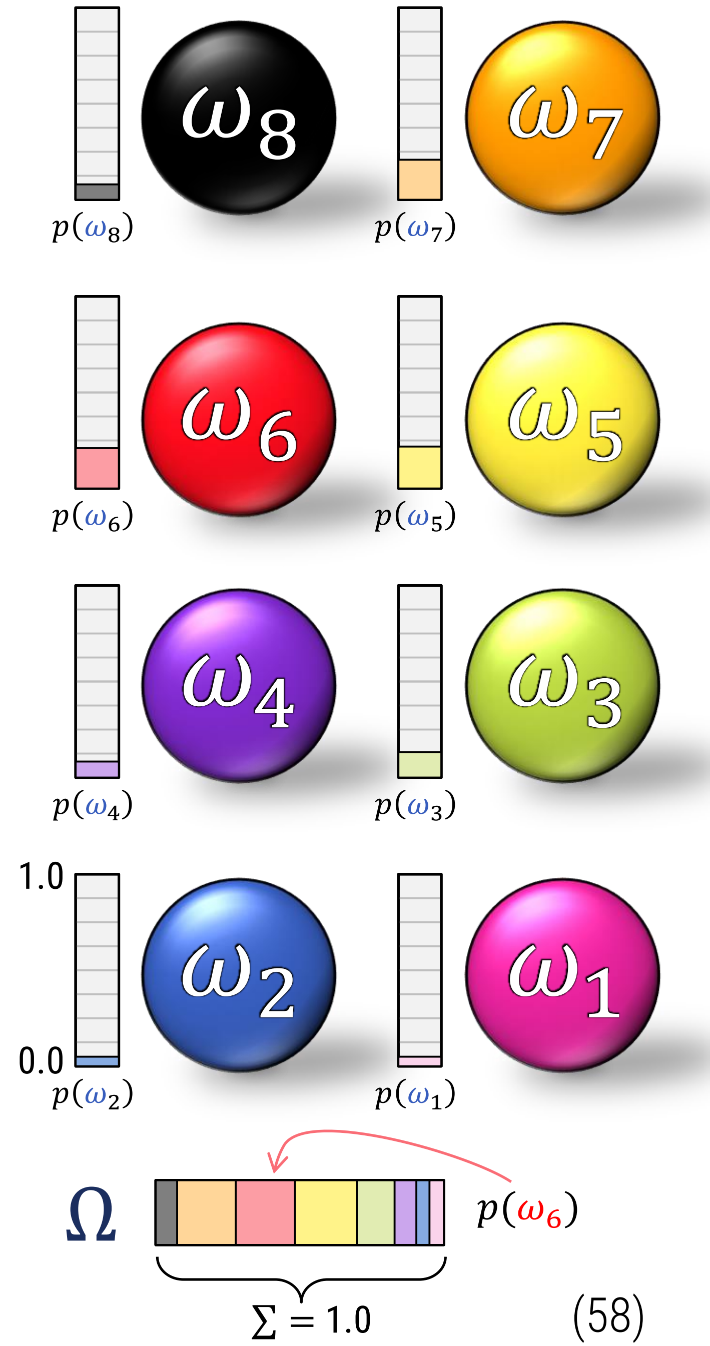
- "*Sample space*" $\Omega = \{\omega_1, \dots, \omega_n\}$
- Outcome $\omega_i \in \Omega$ has probability

$$0 \leq P(\omega_i) \leq 1$$

- The sum of all probabilities is 1

$$\sum_{i=1}^{n} P(\omega_i) = 1$$

- "If we repeat the experiment n times (often), we will observe ω_i roughly $n \cdot P(\omega_i)$ times."



Stochastic Convergence

Probability: *model* of uncertainty

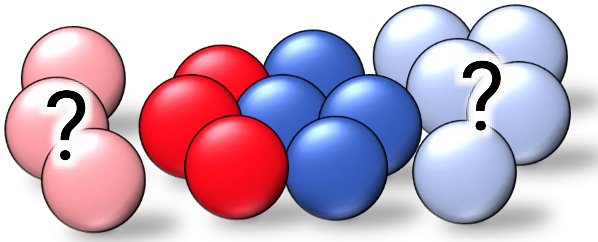
- Motivated by repeating experiment
  - Let $h_n(\omega)$ be the frequency (a random outcome) at which ω was observed in n concrete trials
- $h_n(\omega)$ does not converge to $P(\omega)$ in a classic sense
  - Instead: a "hidden" process makes it unlikely to deviate far
  - Precise: probability of deviation converges to zero

$$\forall \epsilon > 0\colon \lim_{n \to \infty} \left(P(|h_n(\omega) - P(\omega)| > \epsilon) \right) \to 0$$

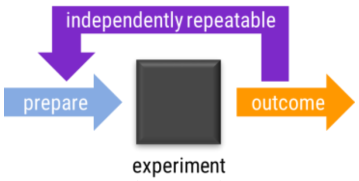
How to Create Knowledge?

(1) Building probabilistic models

- (1a) Theoretical model (class)
  - Prior knowledge (e.g. symmetries)
  - Might contain unknown parameters

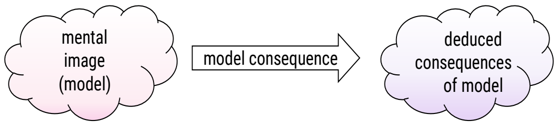


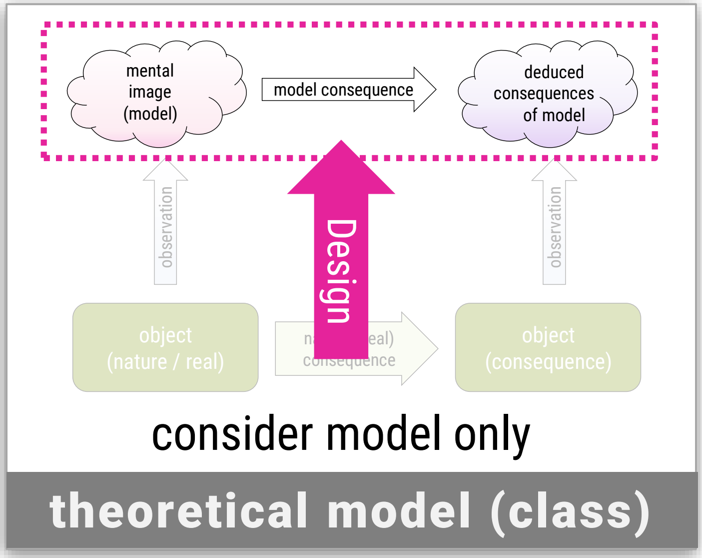
- (1b) Knowledge from experiments
  - Fill in parameter values
  - Statistics / machine learning
  - Prior knowledge always required

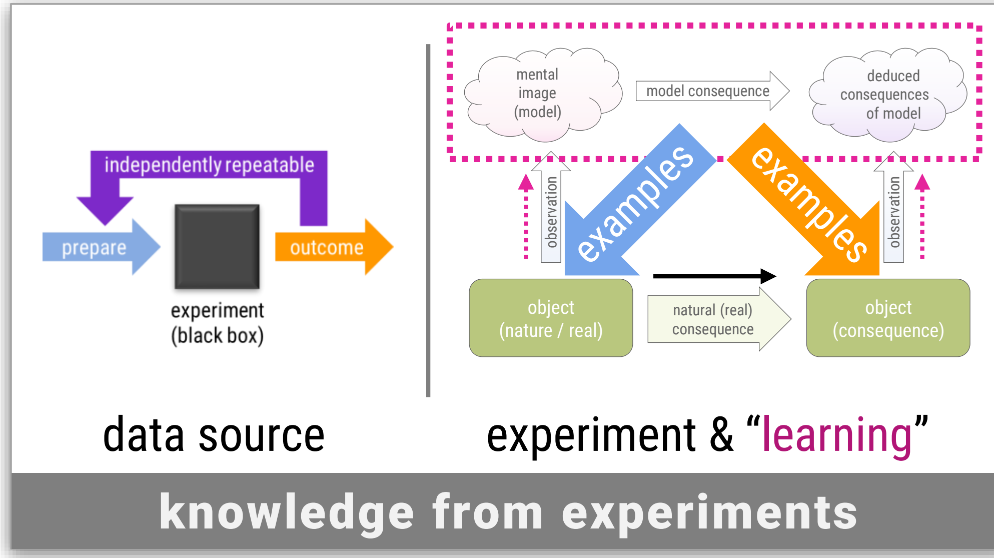


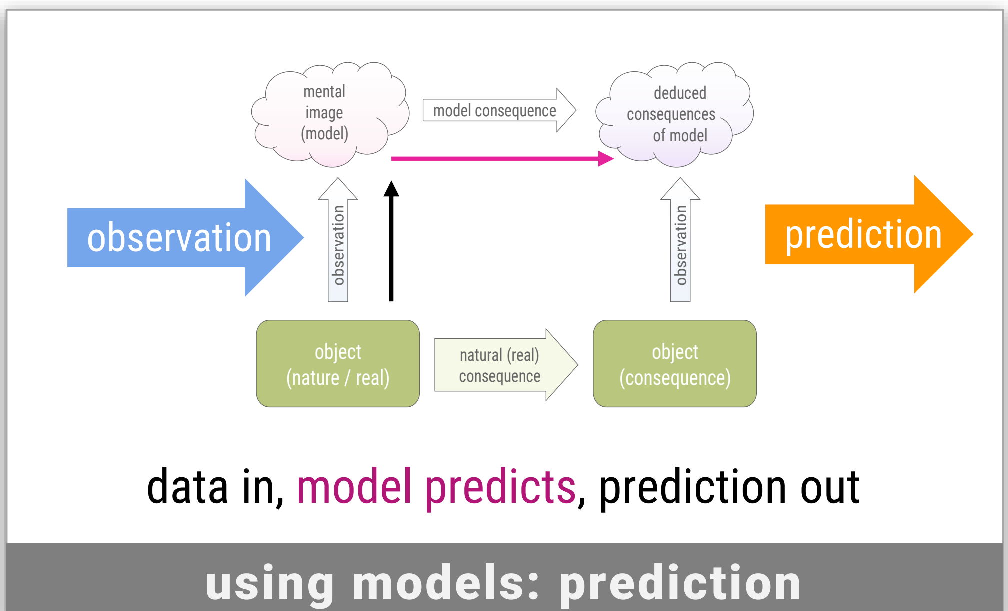
(2) Predictions: using probabilistic models

- New (partial) data / observations
- Infer predictions from models

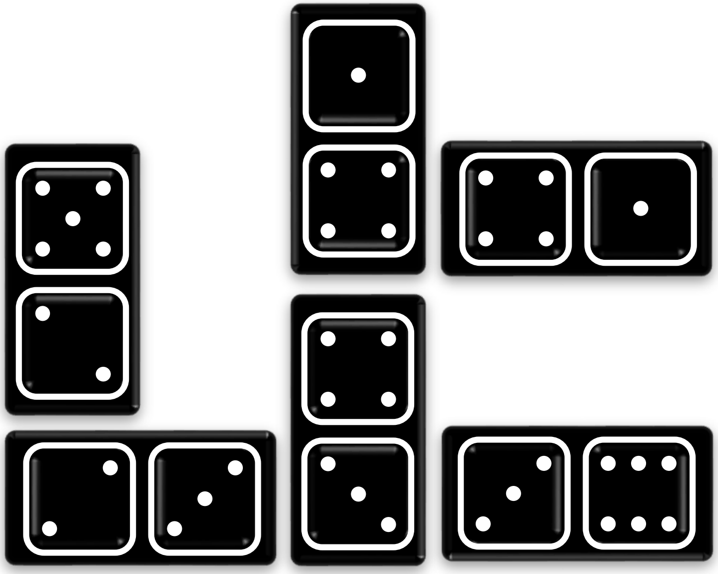



mental
image
(model)
model consequence
deduced
consequences
of model
observation
Design
observation
object
(nature / real)
object
(consequence)
consider model only
theoretical model (class)


independently repeatable
prepare
outcome
experiment
(black box)
mental
image
(model)
model consequence
deduced
consequences
of model
observation
examples
examples
observation
object
(nature / real)
natural (real)
consequence
object
(consequence)
data source
experiment & “learning”
knowledge from experiments


mental
image
(model)
model consequence
deduced
consequences
of model
observation
observation
observation
prediction
object
(nature / real)
natural (real)
consequence
object
(consequence)
data in, model predicts, prediction out
using models: prediction

Summary

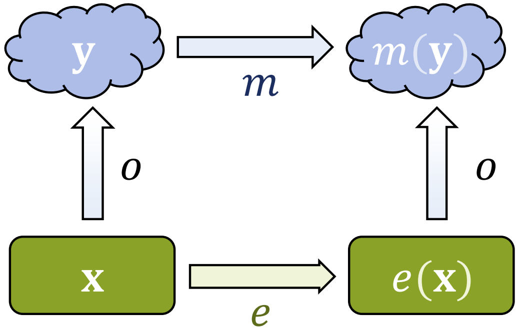


Summary

Gaining knowledge

- Observations → inductive reasoning
- Logical conclusions → deductive reasoning

Algorithmic induction



- Information from observations
- Finite examples → uncertainty
- Statistics models knowledge gain

Will look at probability theory next