Science of Knowledge

Mirco Schönfeld
University of Bayreuth

mirco.schoenfeld@uni-bayreuth.de
@TWlyY29

Day 1: Whisky with ice

https://commons.wikimedia.org/wiki/File:A_Glass_of_Whiskey_on_the_Rocks.jpg

Day 1: Whisky with ice

Day 2: Gin with ice

https://commons.wikimedia.org/wiki/File:Koh_Mak,_Thailand,_Lime_juice_with_ice,_Lemonade,_Limeade.jpg

https://commons.wikimedia.org/wiki/File:Ice_cubes_in_a_glass.jpg

Day 1: Whisky with ice

Day 2: Gin with ice

Day 3: Vodka with ice

Huh?

Drunk on all three days.

What‘s the cause?

Day 1: Whisky with ice

Day 2: Gin with ice

Day 3: Vodka with ice

Somewhere in Between

Epistemology:

- Branch of philosophy concerned with knowledge
- Studies nature, origin, and scope of knowledge
- How does knowledge constitute?

Philosophy of Science:

- Branch of philosophy concerned with science
- Studies foundations, methods, and implications of science
- What qualifies as science?

A Priori vs. a Posteriori Knowledge

A priori knowledge

Independent from experience

Mathematics,
deduction from pure reason

“If George V reigned at least four days, then he reigned more than three days.

A posteriori knowledge

Depends on empirical evidence

Most fields of science
Personal knowledge

“George V reigned from 1910 to 1936.

Techniques of Knowledge Aquisition

Science or not?

In the 19th and 20th century: fundamental upheaval of scientific certainties

Theory of relativity:

Two events appearing simultaneously for one observer aren‘t necessarily appearing so for another.

Quantum Theory:

Light behaves like a wave and like particles *at the same time*

A strong need for philosophic reflection about what is science
and how science can be distinguished from non-science

Nicola, Ubaldo. "Bildatlas Philosophie." Die abendländische Ideengeschichte in Bildern, Berlin (2007).

Verification

Validity of hypotheses is inferred via observations and experiences

Additional proofs further support theories

Philosophers of the Wiener Kreis considered verfication as an important scientific tool:

- Repeated experiments create validity
- Emphasized induction as a scientific tool

Induction

Proposed by Francis Bacon as a scientific method in 1620

Repeated observations are generalized to theories

Universal statements are inferred from singular statements

A form of cognition *a posteriori*

$$H(a_1), H(a_2), \dots, H(a_n) \Rightarrow \forall x: H(x)$$

Induction

https://homepage.univie.ac.at/christian.damboeck/vo14/

https://commons.wikimedia.org/wiki/File:Black_Swan_at_Martin_Mere.JPG https://commons.wikimedia.org/wiki/File:Swan_portrait.jpg https://commons.wikimedia.org/wiki/File:Swan_September_2015-1.jpg
https://commons.wikimedia.org/wiki/File:Mute_swan_(Cygnus_olor)_looking_for_food_in_waves,_Windermere,_England.jpg https://commons.wikimedia.org/wiki/File:Mute_swan_Vrhnika.jpg https://commons.wikimedia.org/wiki/File:CygneVaires.jpg

Problem of Induction

„All swans are white“

How many swans did you see?

And are there exceptions?

No chain of reasoning from observations to inference

In a logical sense: no certainty or universal validity possible

except, of course, complete induction: you have investigated ***all*** phenomena

Formulated in *A Treatise of Human Nature* by David Hume in 1739

Falsification

One single counter-instance is enough to proof a theory wrong

e.g. seeing a black swan

Karl Popper:

Only falsifiable theories should be pursued

According to Popper:

Falsification doesn‘t have to happen

Falsification must logically be possible

Karl Popper (1902-1994)

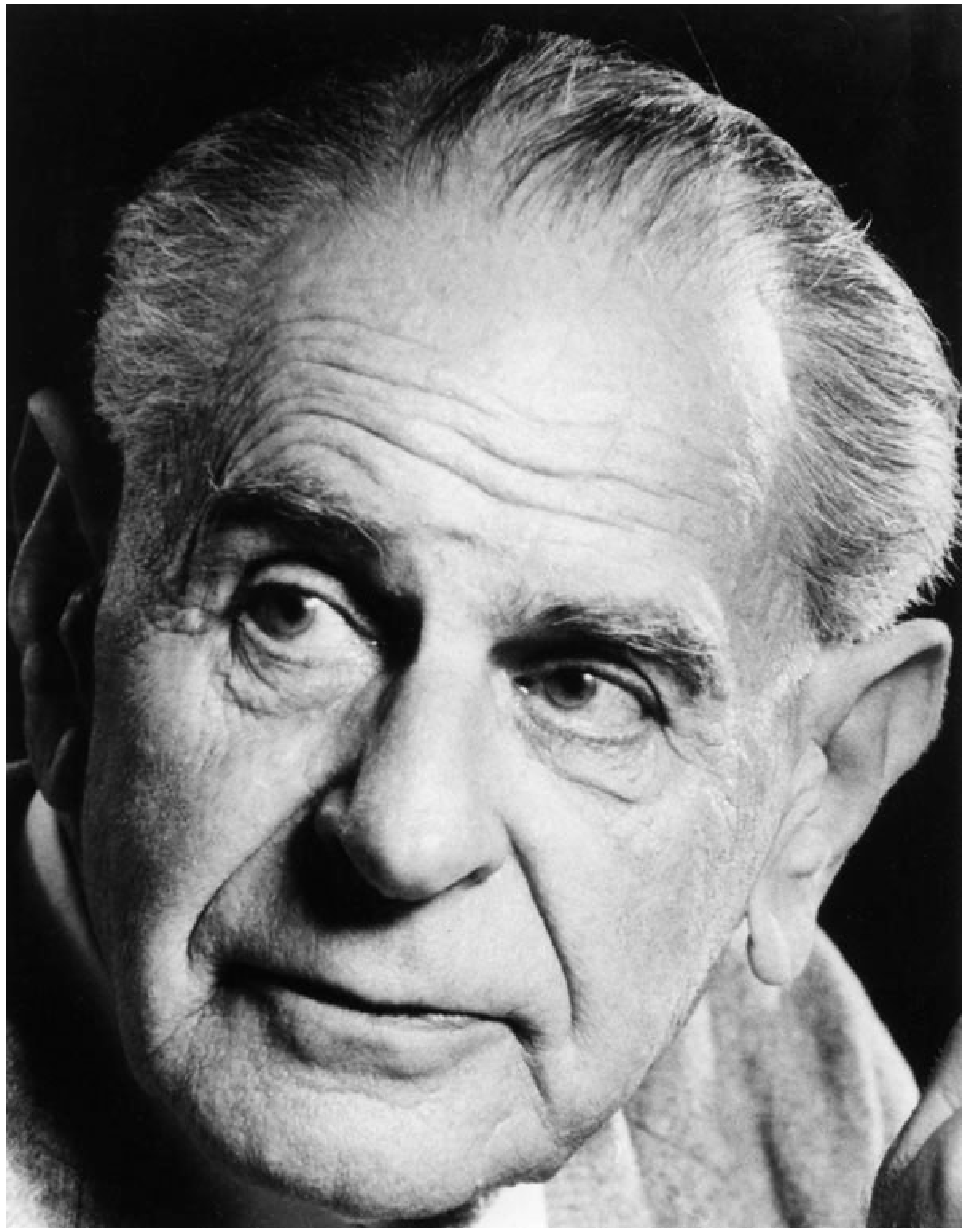
Scientific experimentation is not carried out to verifying or establishing the truth

Popper put a special emphasis on the impor tance of critical spirit to science

False theories can only be eliminated by critical thought

Best theory has the highest level of explanatory force and predictive power

Only logical technique: deductive testing of theories



https://plato.stanford.edu/entries/popper/
https://commons.wikimedia.org/wiki/File:Karl_Popper.jpg

Deduction

Aristotle (384-322 BC)

Axioms are given, theorems are derived

Tool: Deductive Reasoning

1. All men are mortal (*first premise*)
2. Socrates is a man (*second premise*)
3. Therefore, Socrates is mortal (*conclusion*)

No uncertainty:

If all premises are true, the terms are clear,
and the rules of deductive logic are followed,
then conclusions are *necessarily true*.

A form of knowledge *a priori*

Deductive Testing According to Karl Popper

1. Formal testing
   Testing internal consistency

2. Semi-formal
   Investigating the logical form of the theory

3. Comparing
   Does the new theory constitute an advance upon existing ones?

4. Empirical application of conclusions derived from theory
   Positive results never verify a theory.
   Negative results prove that the theory can‘t be completely correct

https://plato.stanford.edu/entries/popper/

Growth of Knowledge (according to Popper)

Design hypotheses such they can be falsified

Rigorosuly try to disprove hypotheses

If succeeded, i.e. a hypotheses is disproved, we learnt something

Repeated failure, i.e. a hypotheses is not disproved in many experiments, doesn‘t mean a theory holds. But it can be accepted at first

All knowledge is provisional, conjectural, hypothetical

https://plato.stanford.edu/entries/popper

But…

Science is based on induction to a great extent

Examples:

- SARS 2002, victim‘s living conditions helped identifying the host animal
- Childbed Fever: disinfection stopped the spread of „cadaveric poison“

Probability

Probability

Science does not produce absolute truth

But rather something that has a certain *reach*, something that is *probably* true

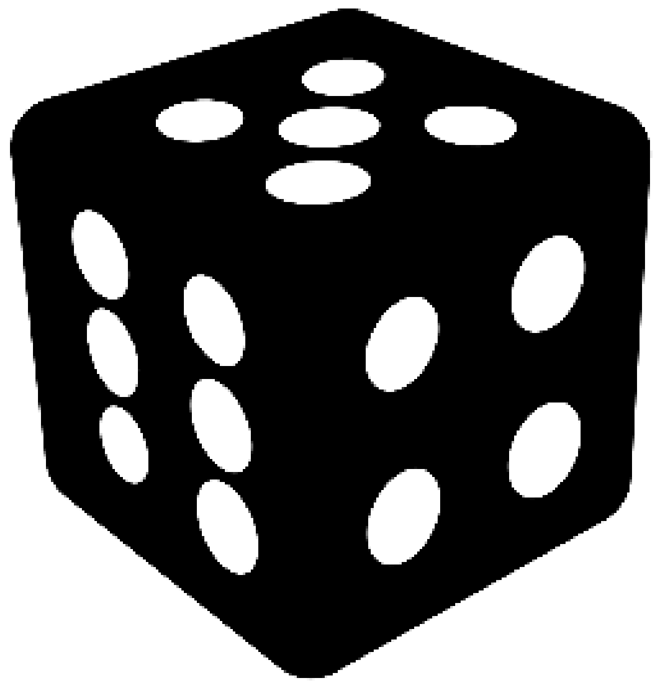
Probabilities of events are numbers between 0 and 1

0 1

Impossible event

E: „Throw a 7“

$P(E) = 0$



Certain event

E: „Throw a number between 1 and 6“

$P(E) = 1$

https://commons.wikimedia.org/wiki/File:Dice_simple.svg

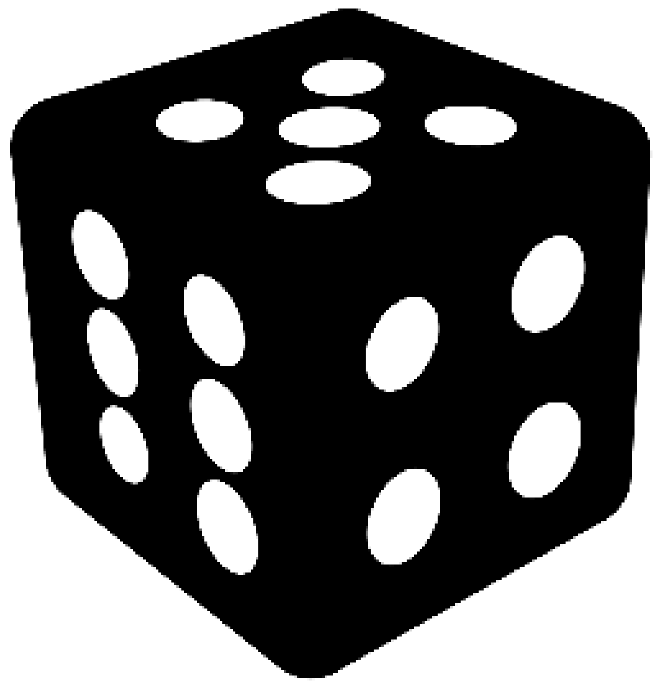
Probability

Probability allows us to draw inferences about the expected frequency of events

$$P = \frac{Number\ of\ favorable\ events}{Number\ of\ possible\ events}$$

0 ——— 1

E: „Throw a 4“, $P(E) = \frac{1}{6}$



E: „Throw 1 or 6“, $P(E) = \frac{2}{6}$

https://commons.wikimedia.org/wiki/File:Dice_simple.svg

Frequentism

Probability expresses *relative frequency* of an event

when an experiment is repeated indefinitely

Probability expresses support of a theory according to the given data

Probability provides additional perspective on classical deductive approach

Frequentism‘s approach

Theory: Smoking causes cancer

We find x percent of smokers having cancer
and y percent of non-smokers having cancer

If we find $x > y$, our theory is supported

Often, conclusions are secured using statistical tools like p-values, standard errors, etc

However, this is all based on the sampling distribution

We know nothing about the *reliability* of the theory

Bayesianism

Thomas Bayes (1701 -1761)

Main idea:
Incorporate the degree of belief in a hypothesis into inferences

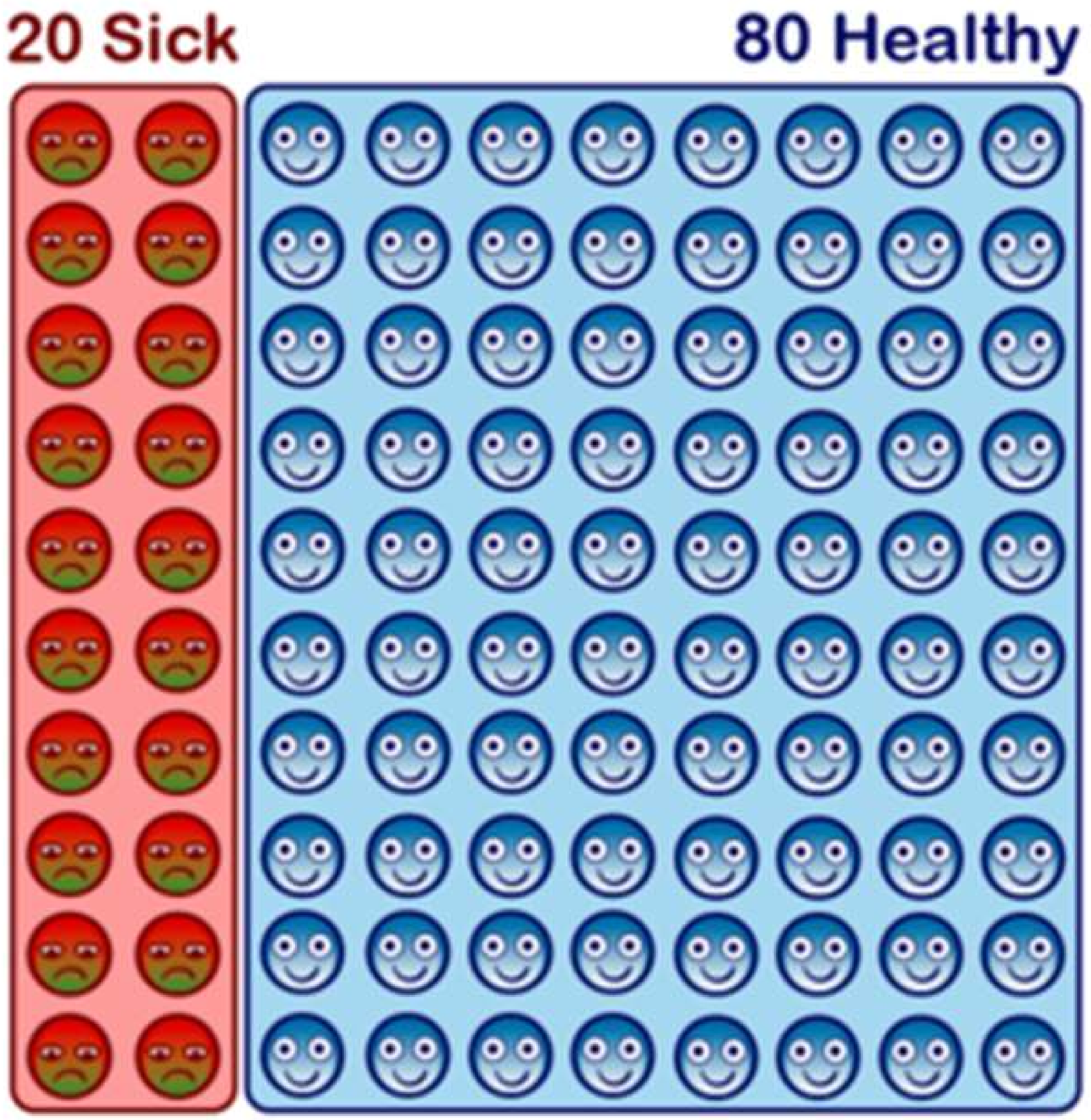
Probabilities also include our own knowledge via *subjective probability*

This adds structure and a degree of confirmation to our hypothesis

Let‘s us encode expert knowledge into the formulation of probabilities

Conditional Probability

Consider a population of 100 students. 20 have Diseasitis and 80 don‘t

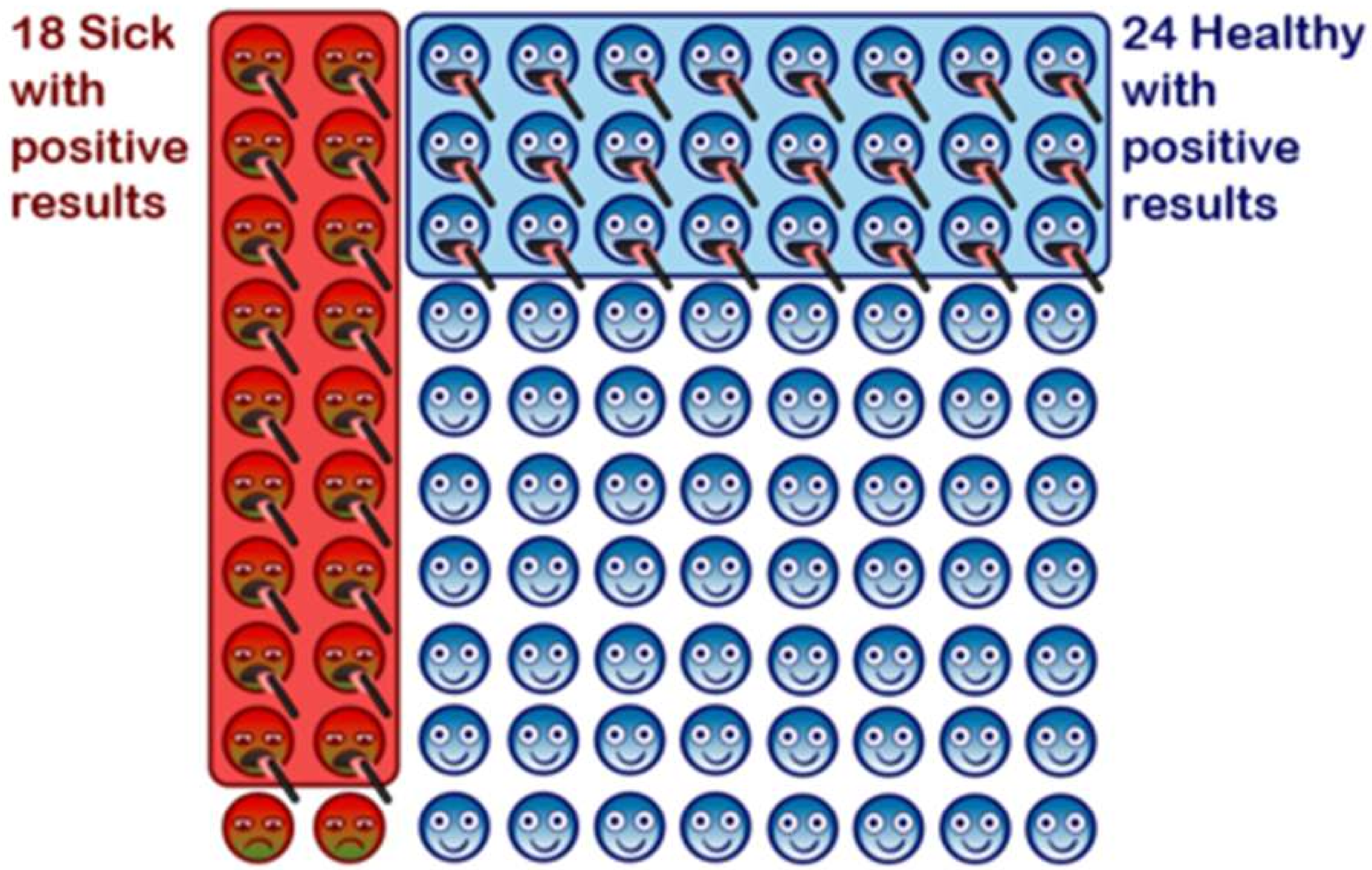


http://arbital.com/p/bayes_rule_guide

Conditional Probability

There is a test for Diseasitis: 90% of sick people test positive, 30% of healthy people too

$P(pos|sick) = 0.9$



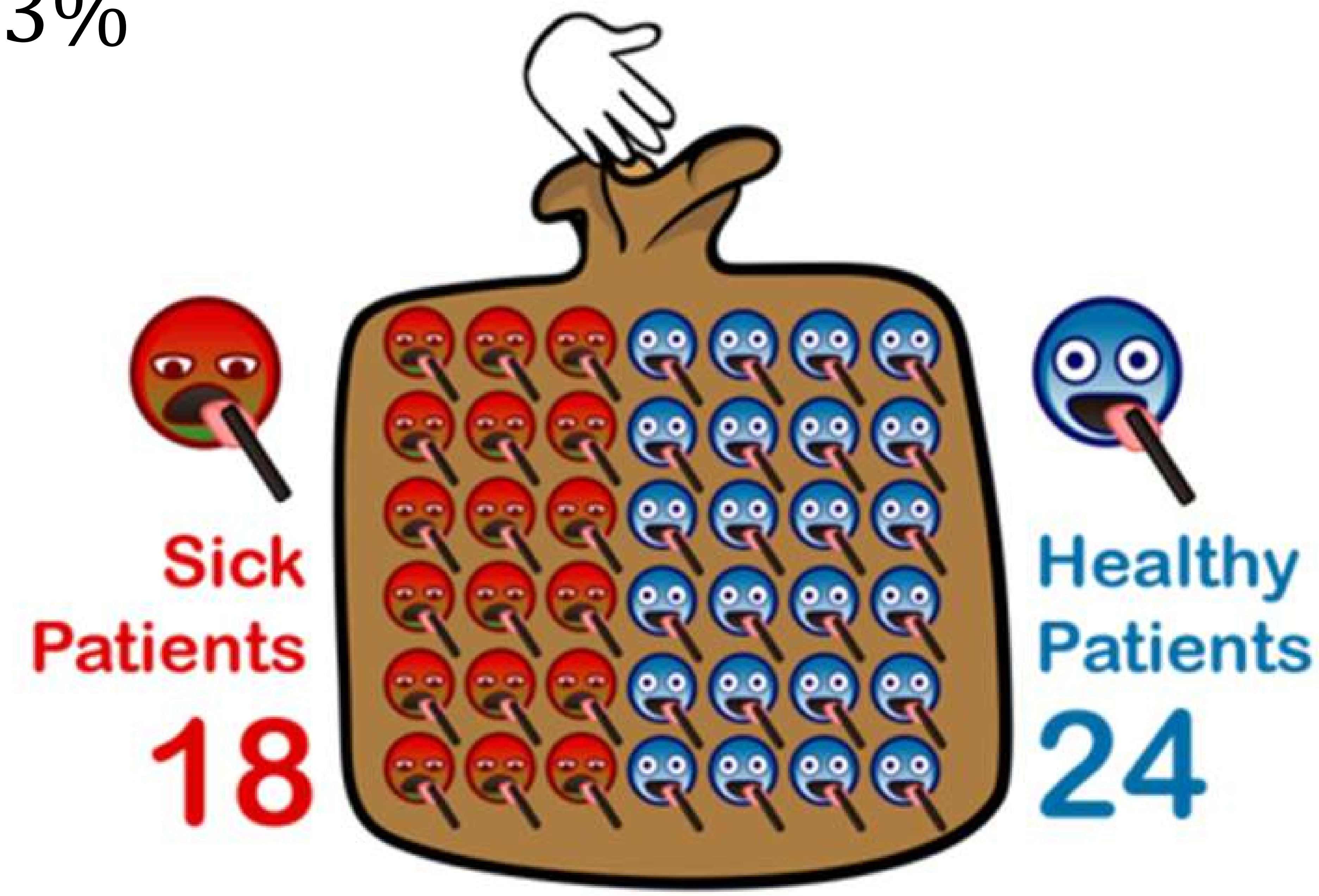
$P(pos|\neg sick) = 0.3$

http://arbital.com/p/bayes_rule_guide

Conditional Probability

If your test result is positive, how likely are you infected?

$$P(sick|pos) = \frac{18}{42} = \frac{3}{7} \approx 43\%$$



Counter intuitive?
The test detects 90% of sick people!

But still below 50% to have Diseasitis?

Well, the test provides *some* evidence:
Before, the probability of being sick was 20/100 = 20%, now it is 43%!

$$P(sick|pos) := \frac{P(sick \wedge pos)}{P(pos)} = \frac{18}{18+24}$$

http://arbital.com/p/bayes_rule_guide

Bayesian Priors and Posteriors

$P(H)$:

Prior probability allows us to formulate our assumption of the probability of a hypothesis

$P(H|E)$:

Then, empirical evidence updates the prior probability yielding a **posterior probability** that incorporates both the prior and the empirical observation

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

Bayes‘ Rule

$P(H)$: Prior probability

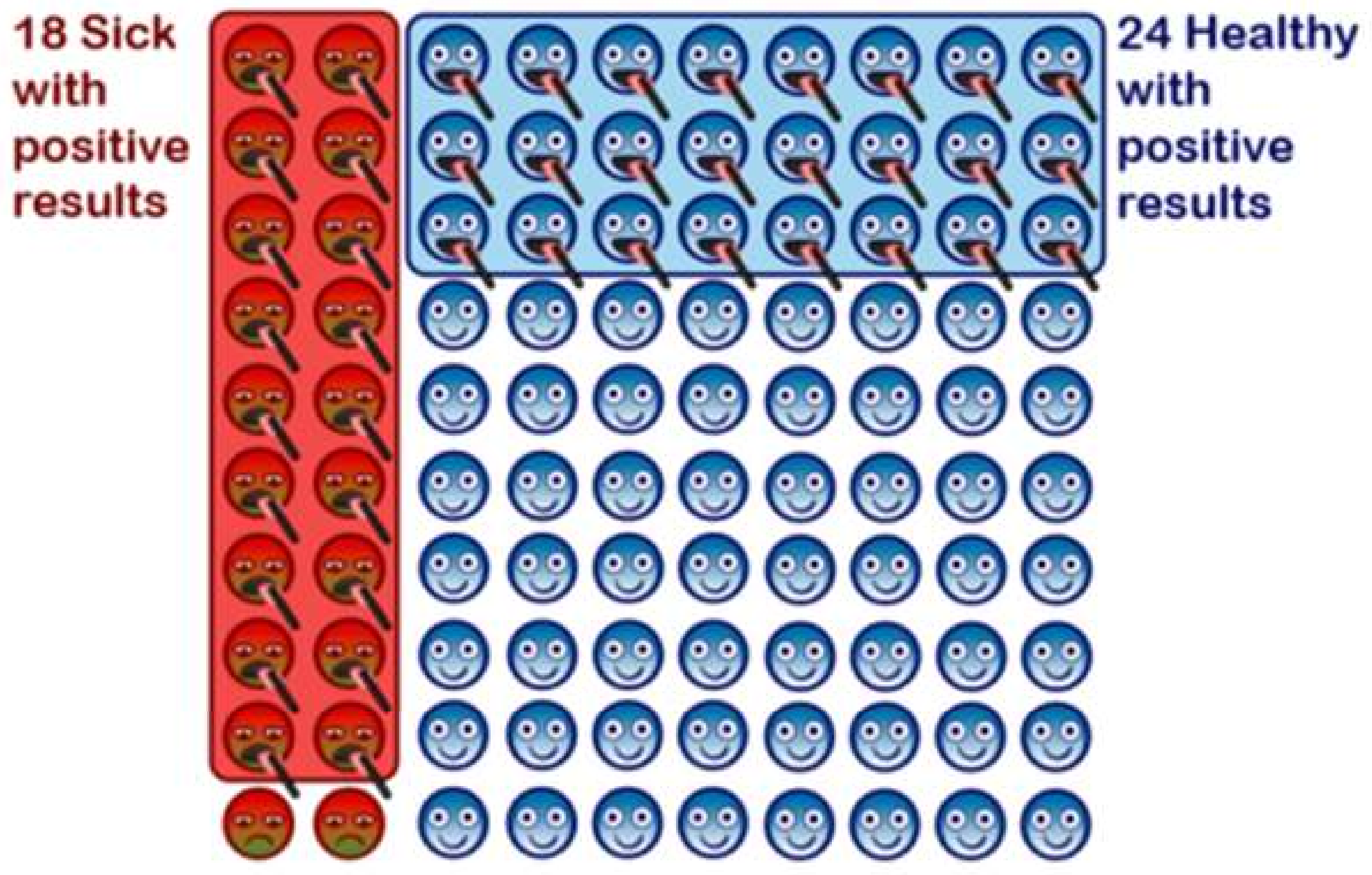
$$P(sick) = \frac{20}{100} = 20\%$$

$P(E)$: Probability of evidence E

$$P(pos) = \frac{42}{100} = 42\%$$

$P(E|H)$: Probability of E under the assumption H

$$P(pos|sick) = \frac{18}{20} = 90\%$$



Posterior probability $$P(H|E) = \frac{P(E|H)P(H)}{P(E)} = \frac{0.9x0.2}{0.42} \approx 0.43$$

http://arbital.com/p/bayes_rule_guide

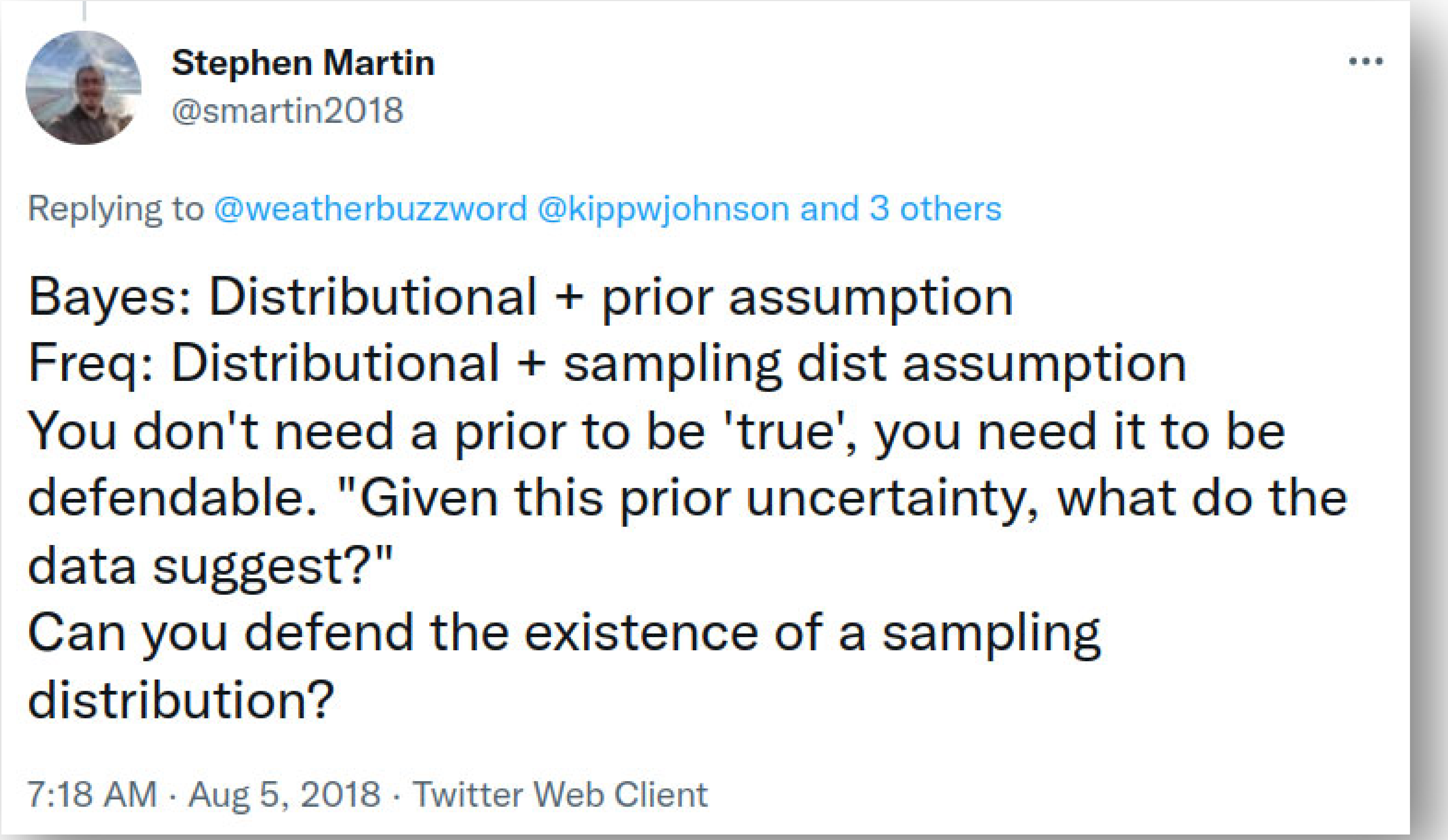
Subjectivity is not Arbitrariness

UNIVERSITÄT BAYREUTH

Main difficulty is the initialization of prior probability

Priors introduce a certain subjectivity

Priors don‘t need to be *true*,
they just need to be *defendable*

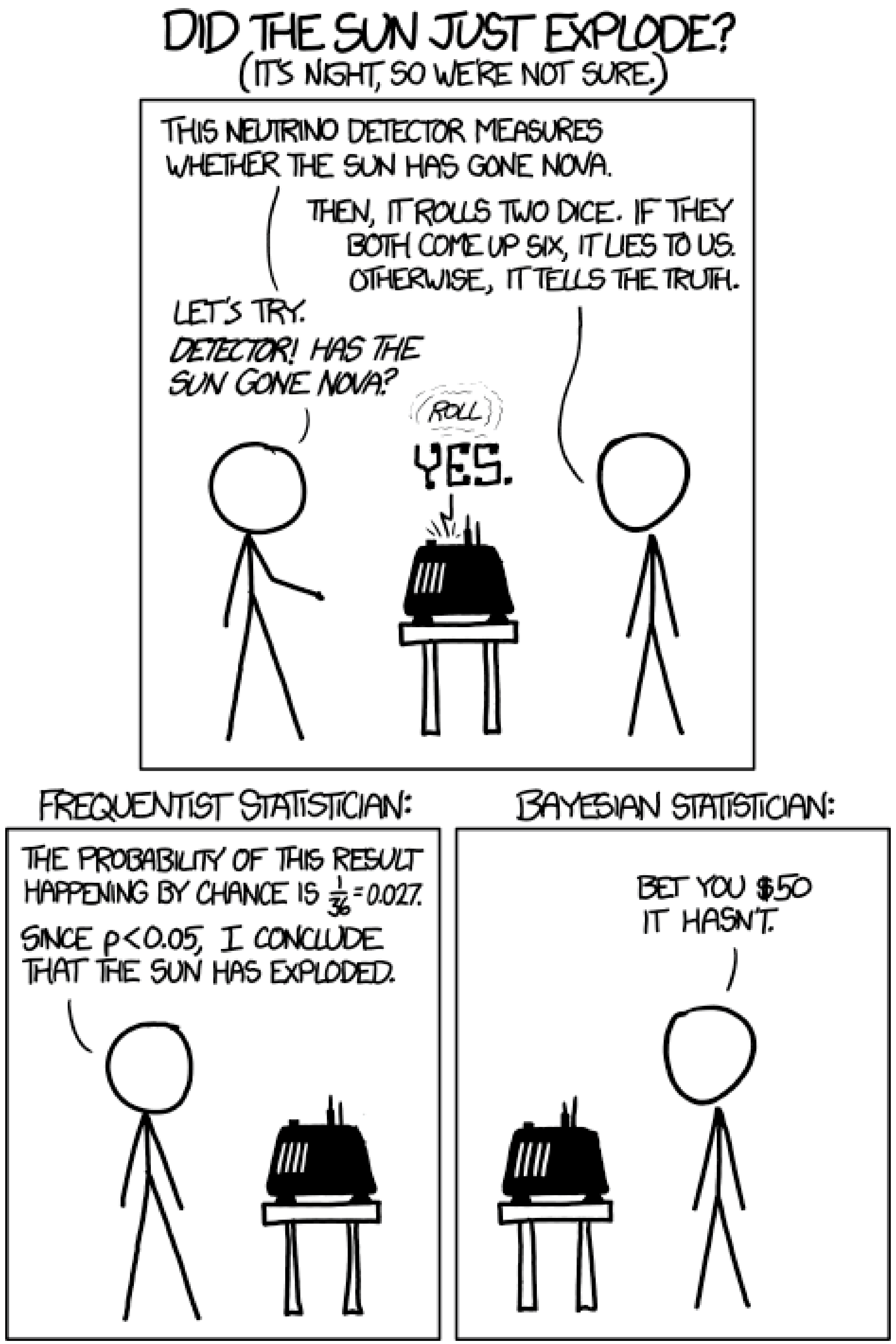


Integrating prior probability is a model for considering current state of knowledge

„Learning“ means updating prior probabilities based on data

https://twitter.com/smartin2018/status/1025974302607388672

Did the sun just explode?



https://xkcd.com/1132/

What is Machine Learning after all?

Statistical Learning is inductive!

Data describe specific observations from the past

Training a model means inferring general rules from the given observations

Statistical Learning is deductive!

Using a learned model to make predictions is a type of deductive inference

However, models fail at generalizing beyond training data

UNIVERSITÄT BAYREUTH

> …the majority of current successes of machine learning boil down to large scale pattern recognition on suitably collected *independent and identically distributed* (i.i.d.) data.

Schölkopf, Bernhard, et al. "Toward causal representation learning." Proceedings of the IEEE 109.5 (2021): 612-634.

Independent and Identically Distributed Data

Central assumptions in Machine Learning:

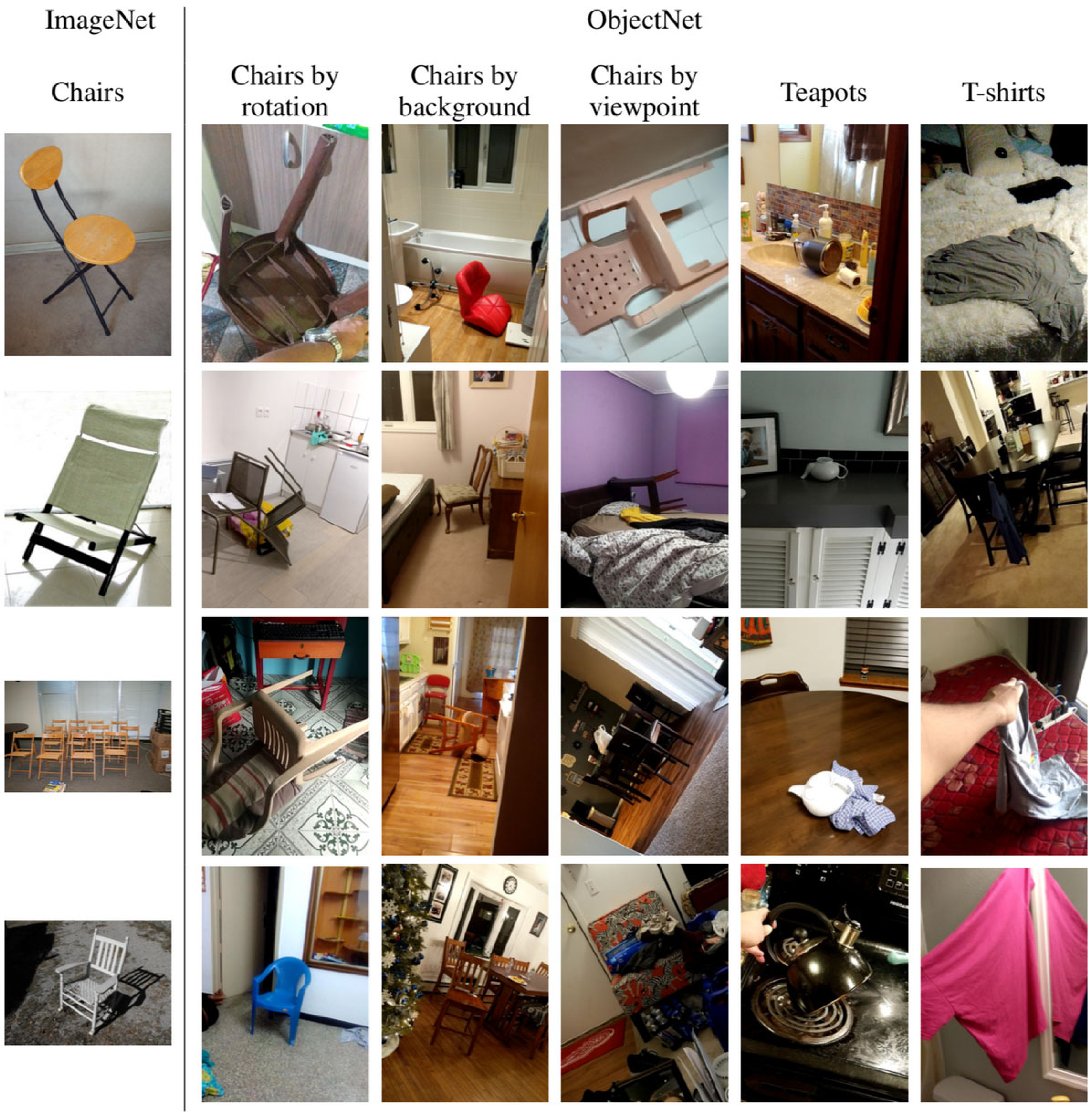
- Observations are not dependent on each other
- Observations have a constant probability of occuring

Consequence:

If the training set is large enough, the ML model will be able to generalize appropriately

Appealing aspects:

- Scalable
- Easy to evaluate

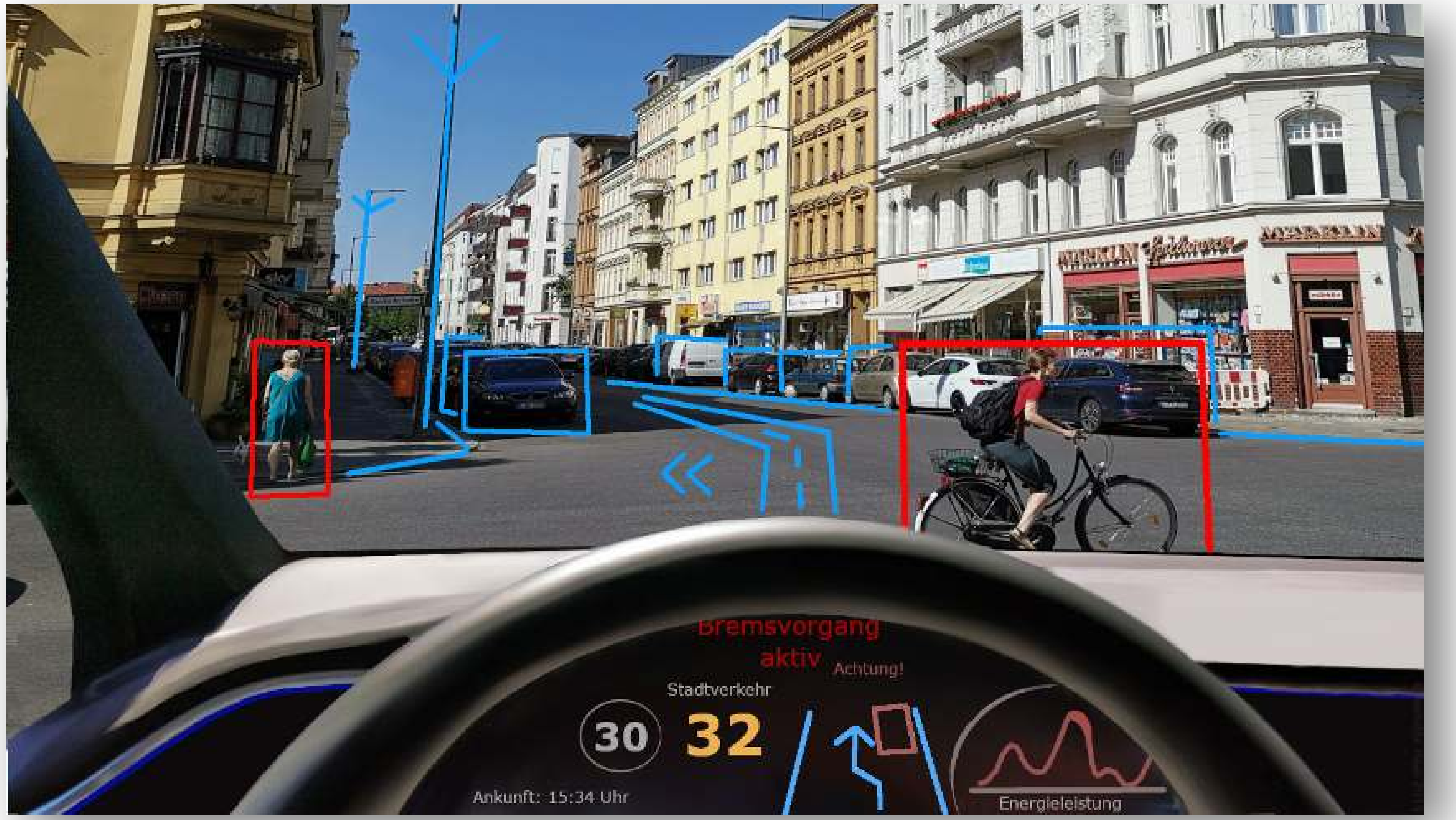


https://objectnet.dev/

Too Much Complexity

At some point, it will become impossible to cover the entire distribution of expected cases by collecting more training data

Especially in contexts of Artificial Intelligence



https://commons.wikimedia.org/wiki/File:Autonomous-driving-Barcelona.jpg

UNIVERSITÄT BAYREUTH

“Generalizing well outside the i.i.d. setting requires learning not mere statistical associations between variables, but an underlying causal model

Schölkopf, Bernhard, et al. "Toward causal representation learning." Proceedings of the IEEE 109.5 (2021): 612-634.

Causal Models

…remain robust when interventions change the statistical distribution of a problem

…will allow a machine to respond to unseen situations

…will allow us to think about counterfactuals

…will be crucial in dealing with adversarial attacks

…will help tackle Machine Learning‘s lack of generalization

https://bdtechtalks.com/2021/03/15/machine-learning-causality/

Rigby

Enough with the wind already

Received April 1

Ever since they installed all those big fans up on the hill it's become even windier. Whose bright idea was that?

I've noticed when they're off, we get a nice calm spell. Please turn them off, at least on weekends. (Word count: 40)

JEFF FORBES
Idaho Falls

ds max • **Guest columns, solicited:** 450 words max • **Guest column**

Thanks.

mirco.schoenfeld@uni-bayreuth.de