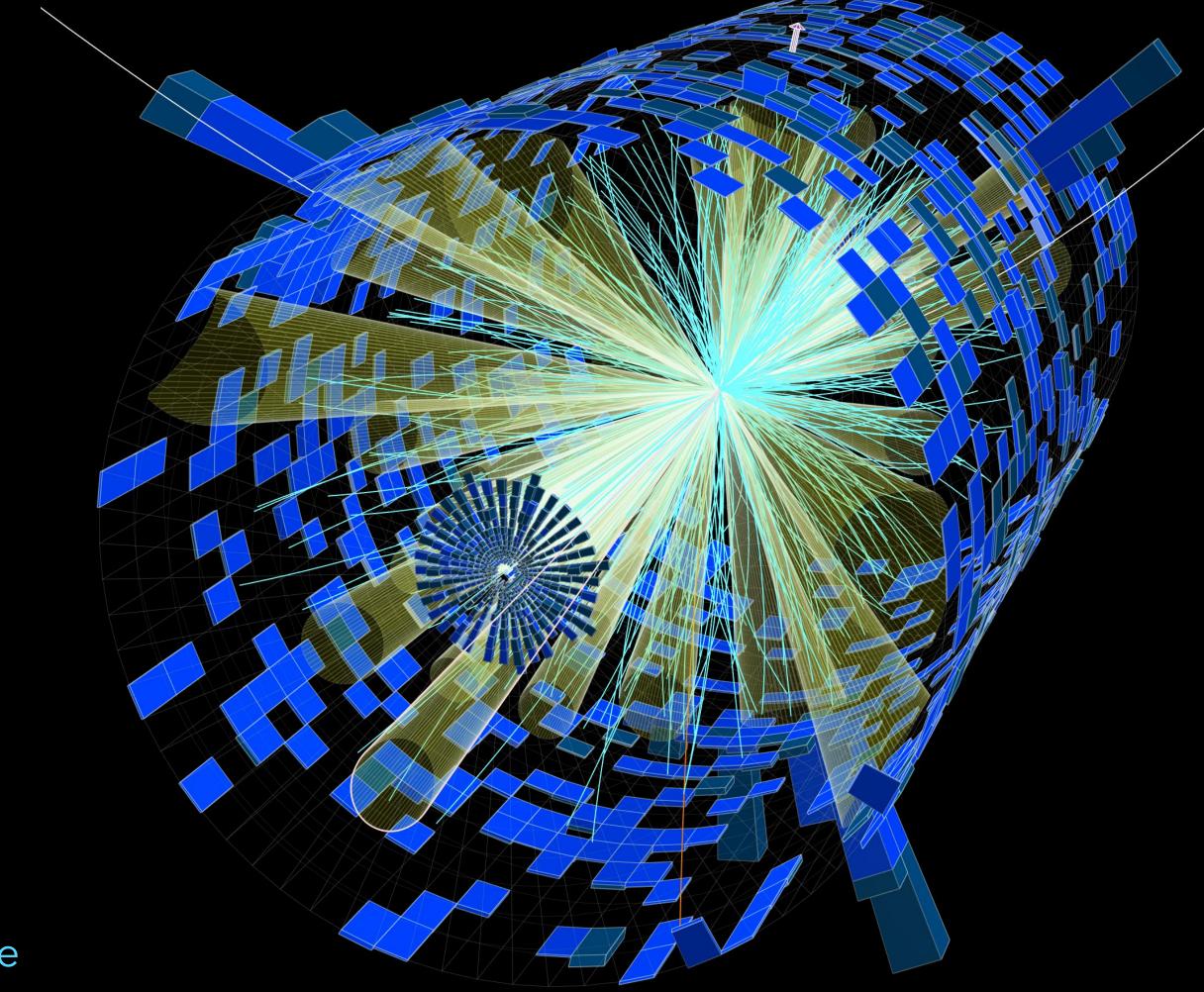
THE INTERSECTION OF

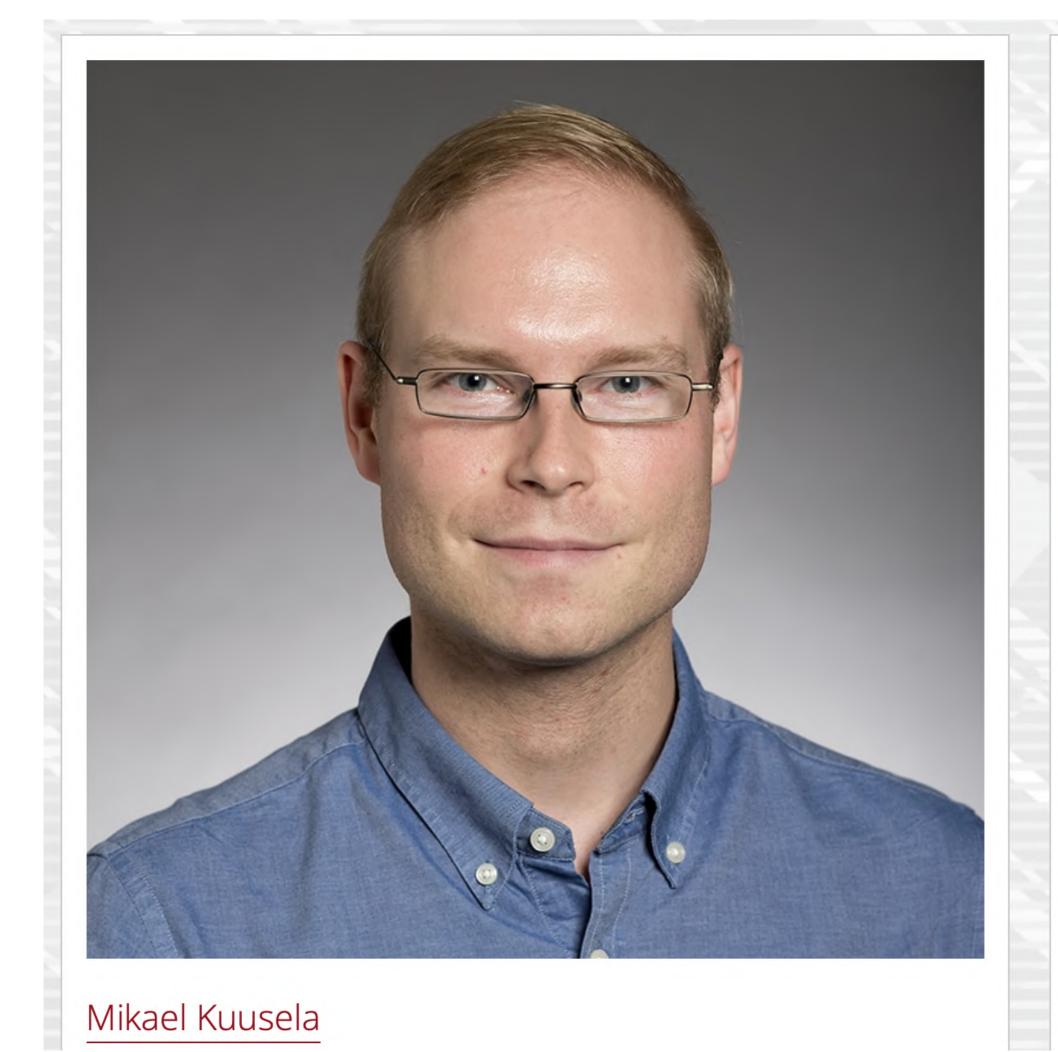
STATISTICS, MACHINE LEARNING, AND THE PHYSICAL SCIENCES

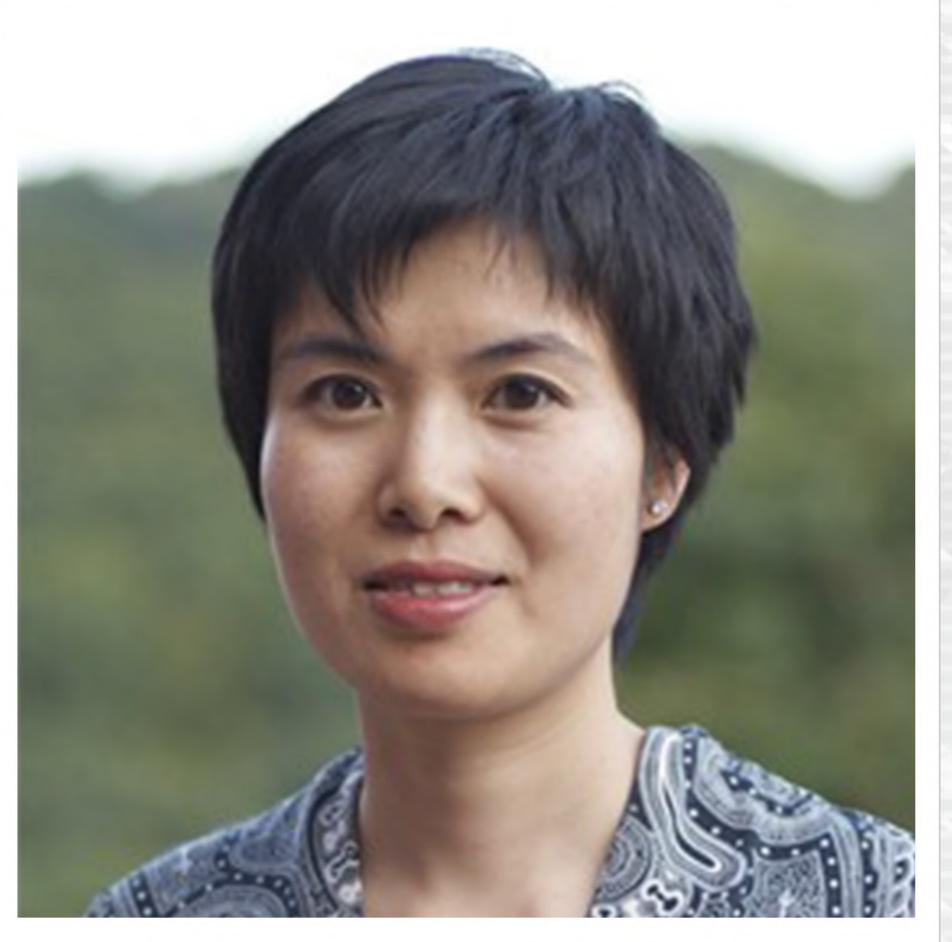




Congratulations!

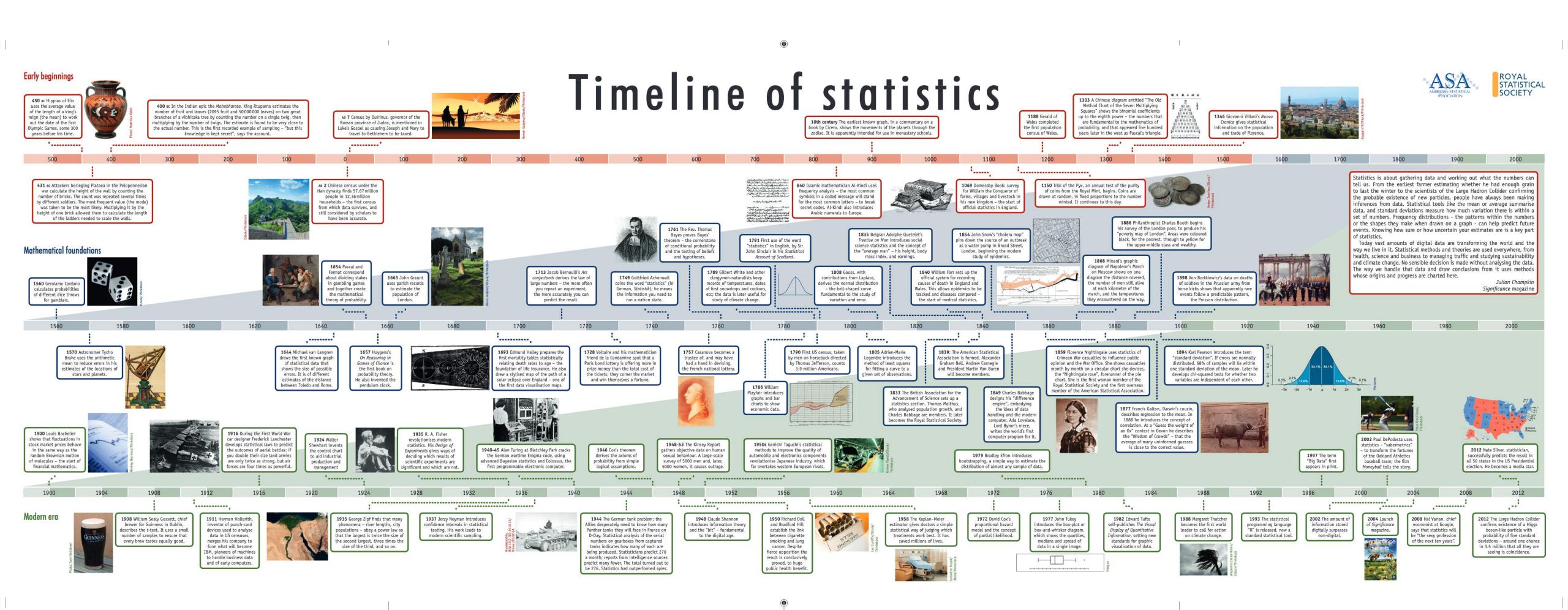
... on the launch of STAMPS@CMU!

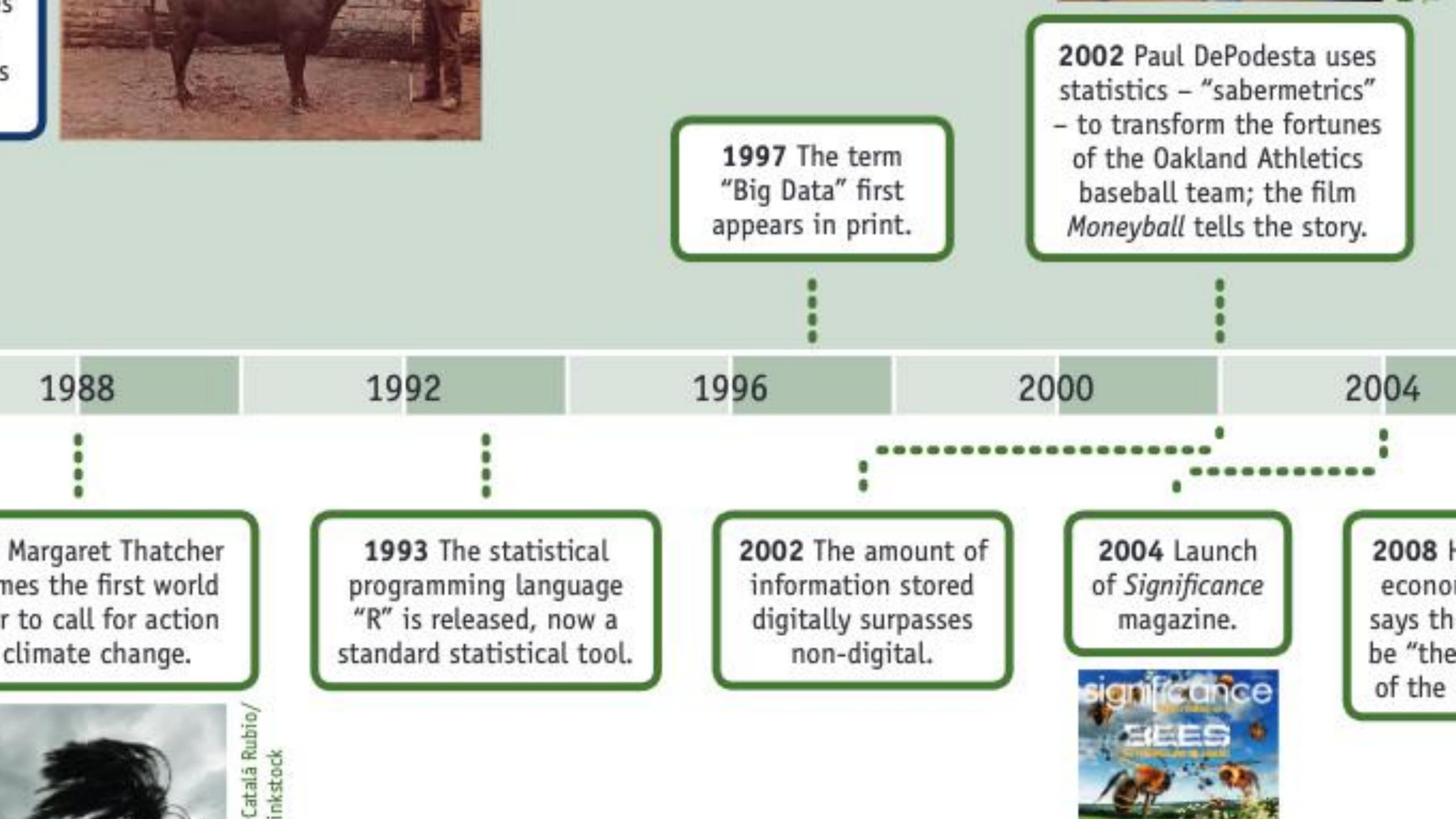


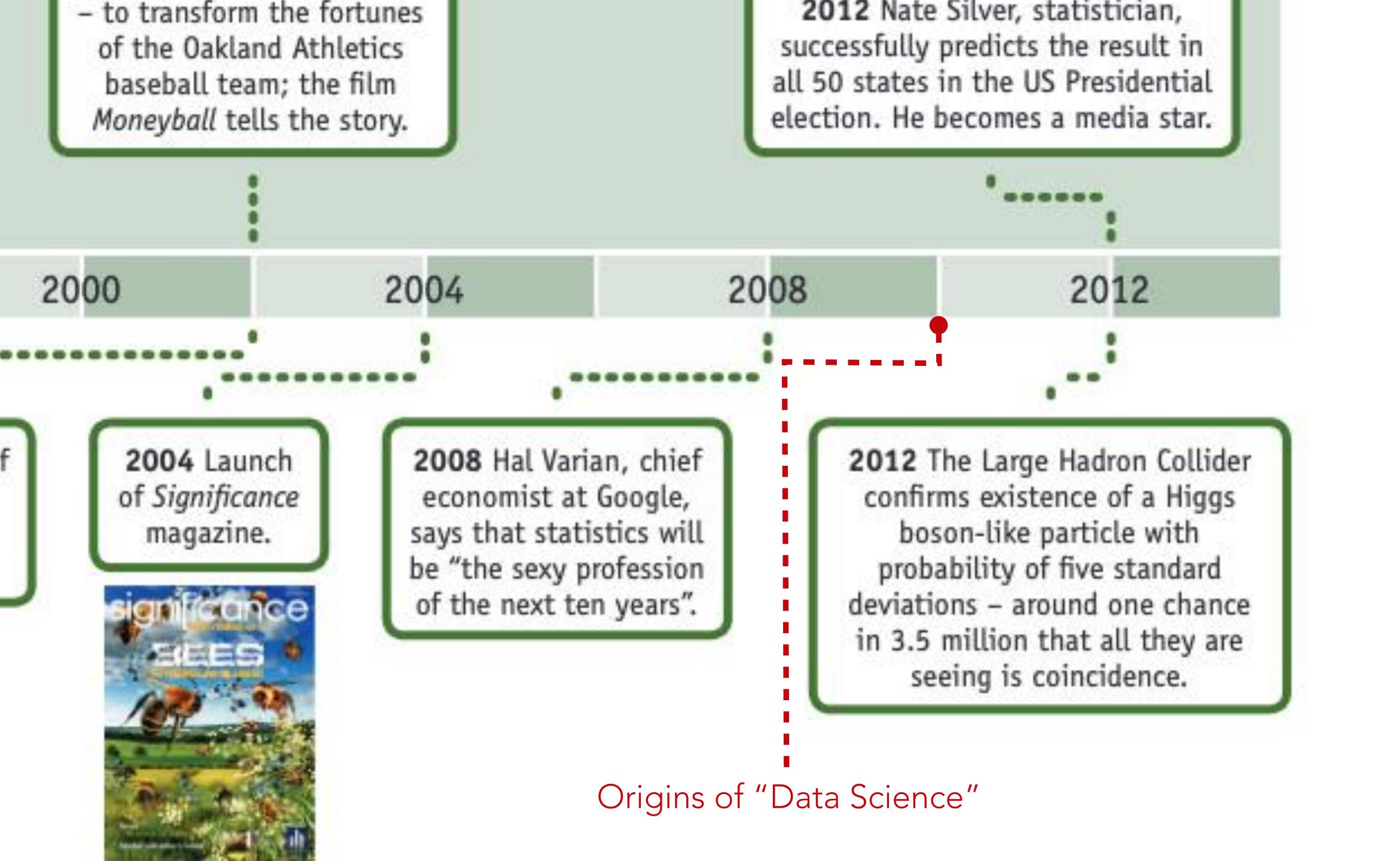


Ann Lee

How did we get here? Big Data and Machine Learning







Statistical Science 2001, Vol. 16, No. 3, 199–231

Statistical Modeling: The Two Cultures

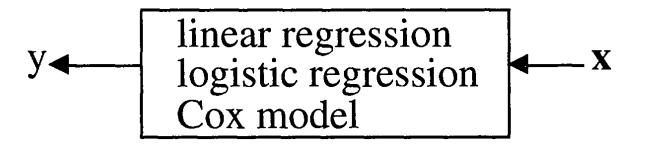
Leo Breiman

Abstract. There are two cultures in the use of statistical modeling to reach conclusions from data. One assumes that the data are generated by a given stochastic data model. The other uses algorithmic models and treats the data mechanism as unknown. The statistical community has been committed to the almost exclusive use of data models. This commitment has led to irrelevant theory, questionable conclusions, and has kept statisticians from working on a large range of interesting current problems. Algorithmic modeling, both in theory and practice, has developed rapidly in fields outside statistics. It can be used both on large complex data sets and as a more accurate and informative alternative to data modeling on smaller data sets. If our goal as a field is to use data to solve problems, then we need to move away from exclusive dependence on data models and adopt a more diverse set of tools.

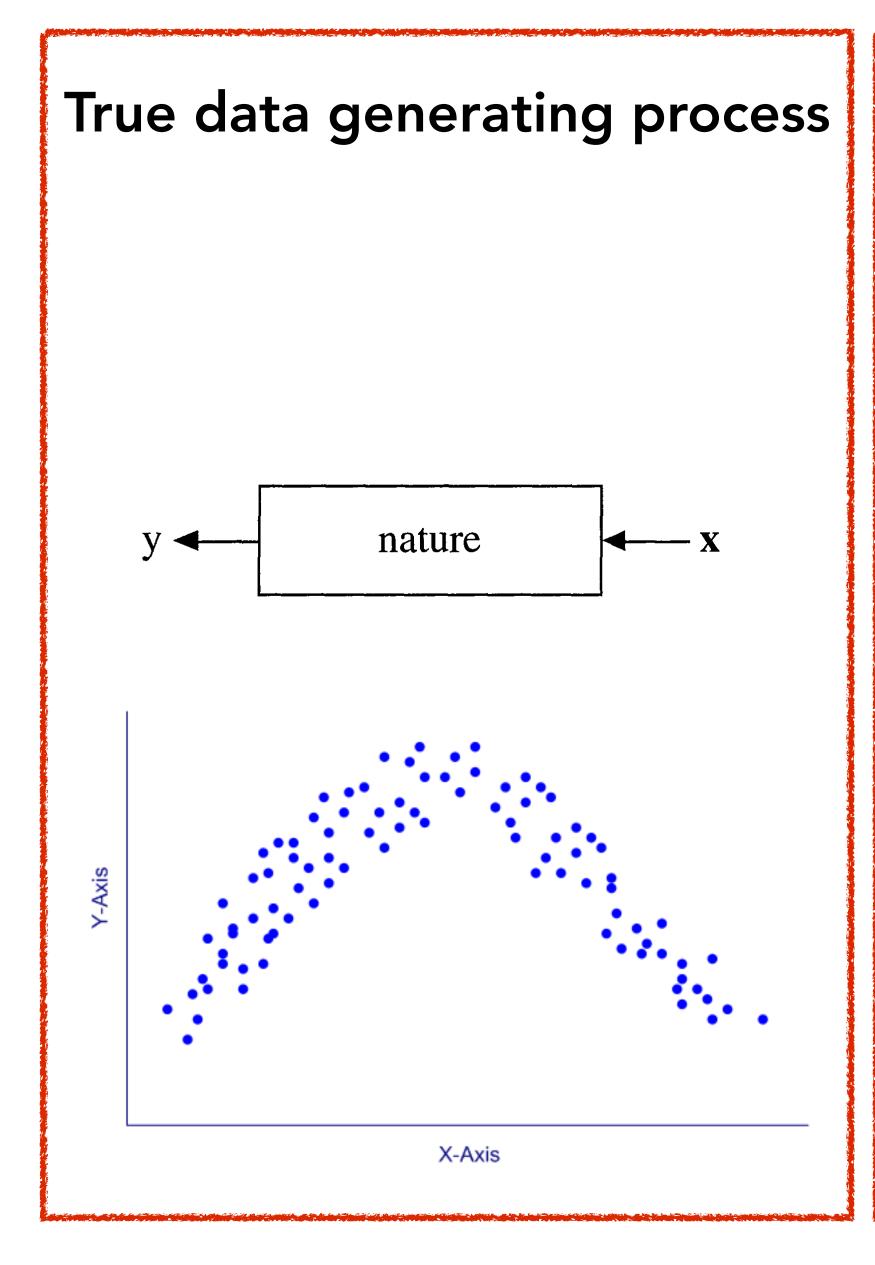
True data generating process

y **nature X**

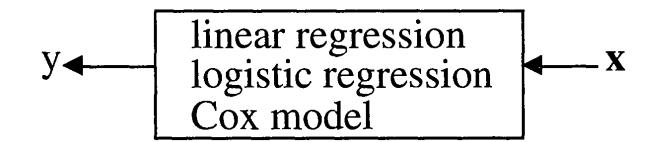
Classical statistical modeling



"Assume data are generated by the following model..."

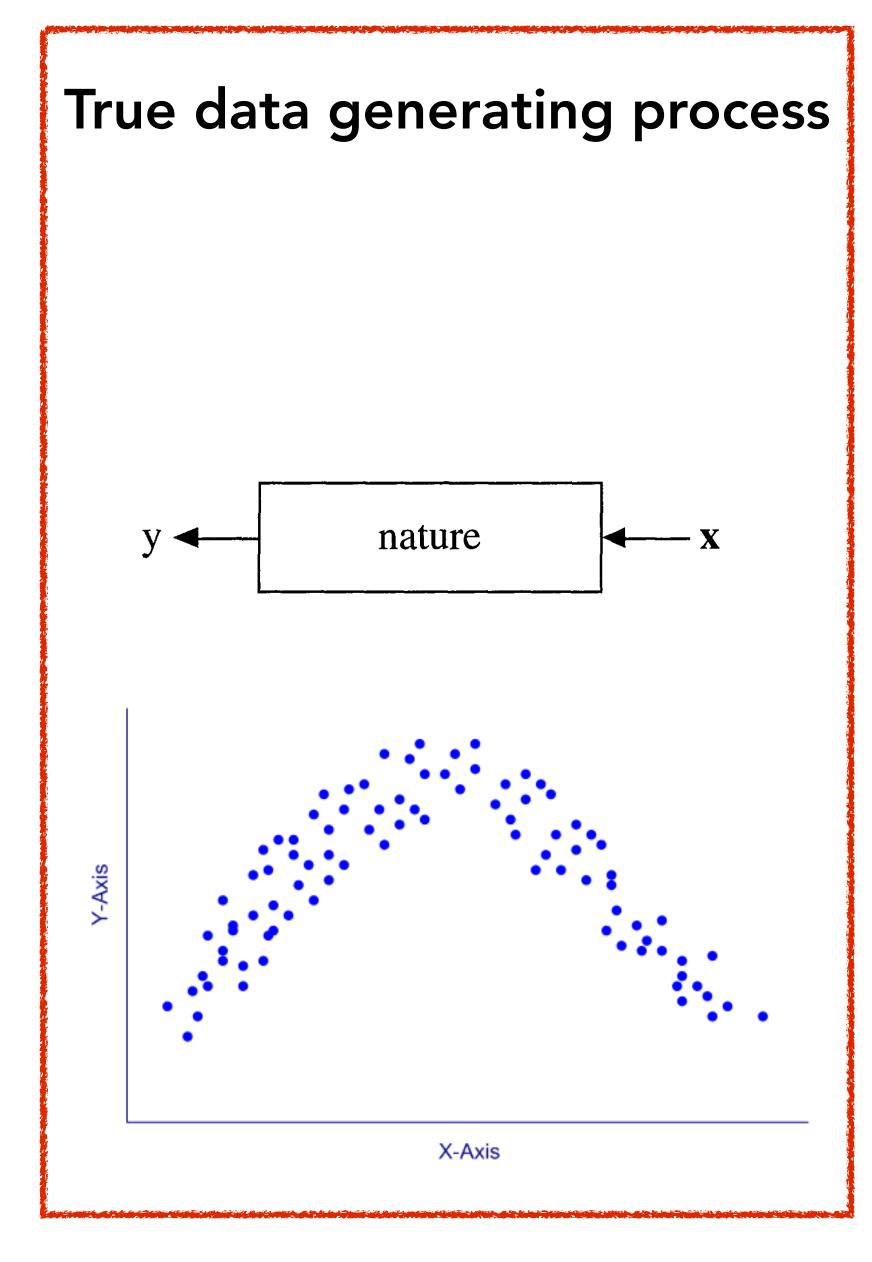


Classical statistical modeling

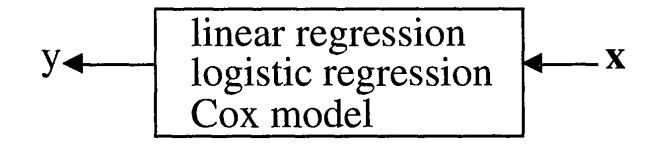


"Assume data are generated by the following model..."

"This enterprise has at its heart the belief that a statistician, by imagination and by looking at the data, can invent a reasonably good parametric class of models for a complex mechanism devised by nature."



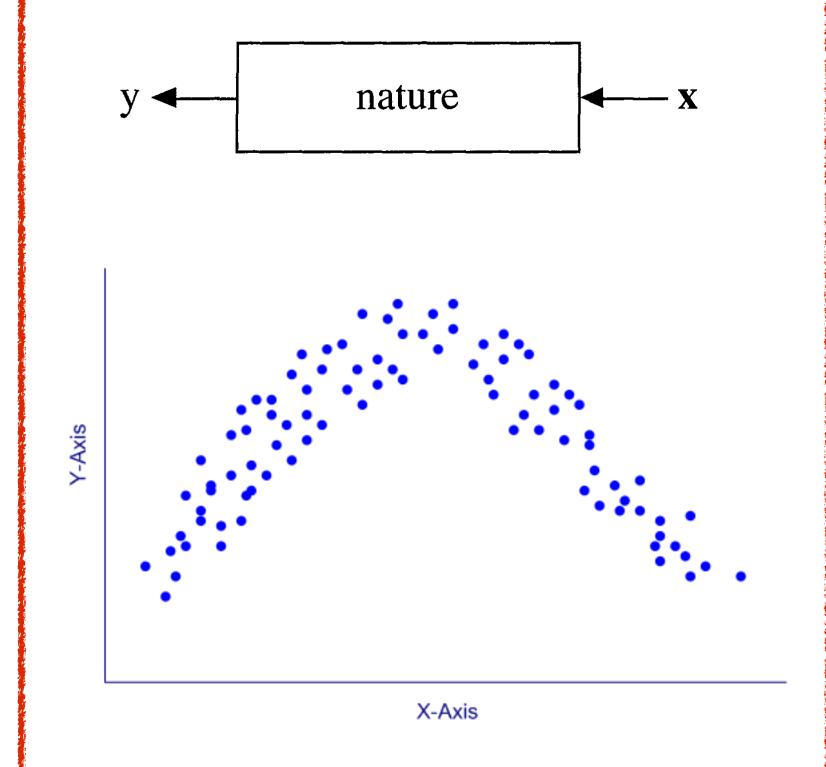
Classical statistical modeling



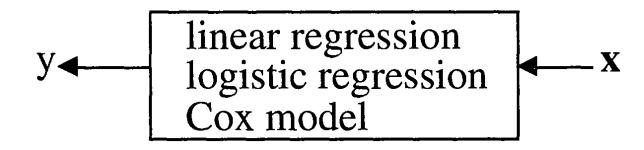
"Assume data are generated by the following model..."

"This enterprise has at its heart the belief that a statistician physicist, by imagination and by looking at the data, can invent a reasonably good parametric class of models for a complex mechanism devised by nature."

True data generating process

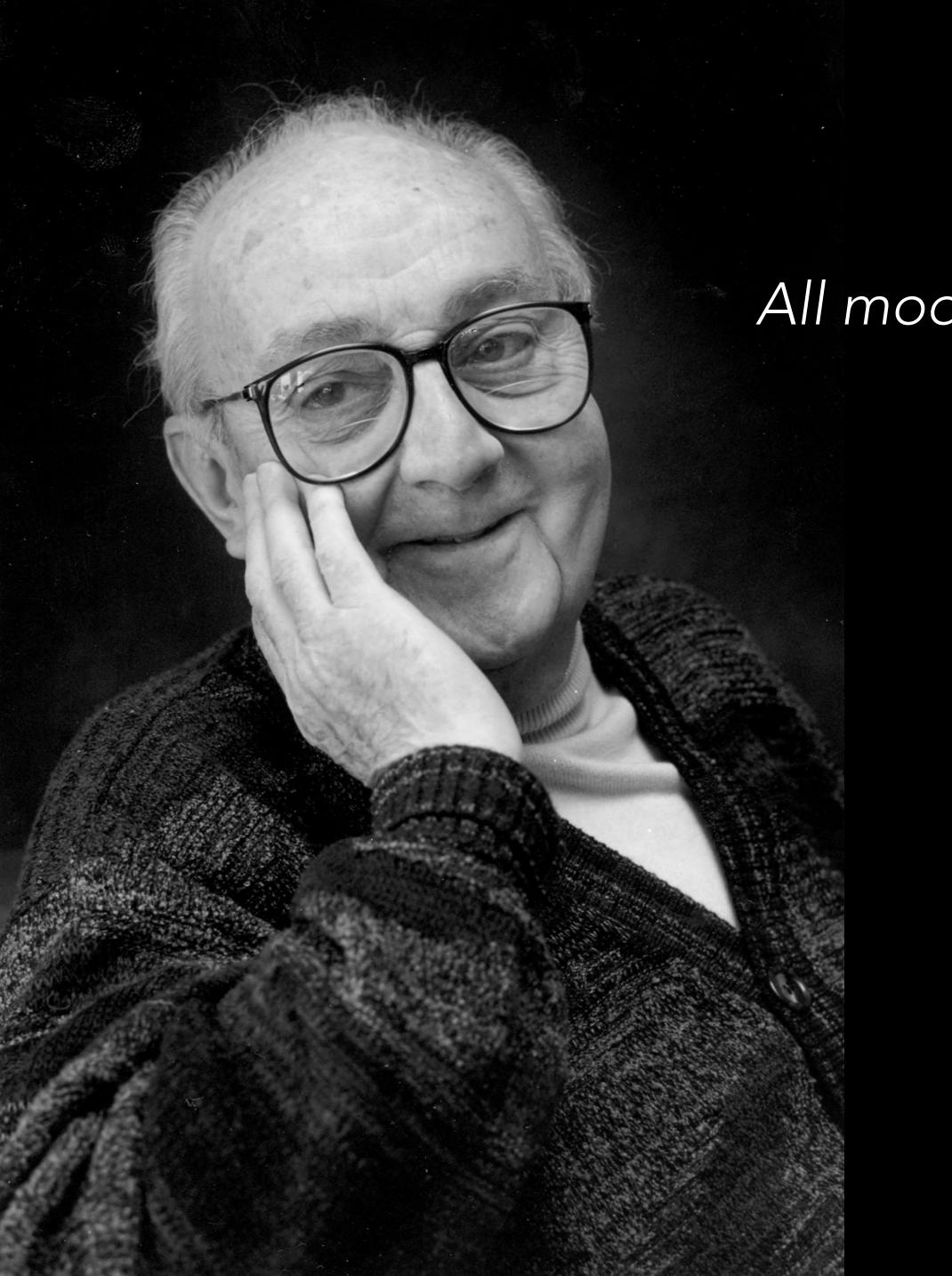


Classical statistical modeling



"Assume data are generated by the following model..."





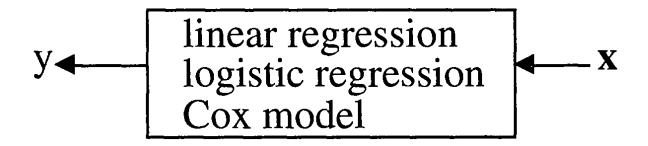
All models are wrong, but some are are useful.

-GEORGE BOX, UW-MADISON

True data generating process

y **nature X**

Classical statistical modeling

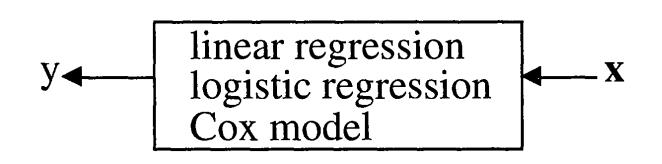


"Assume data are generated by the following model..."

True data generating process

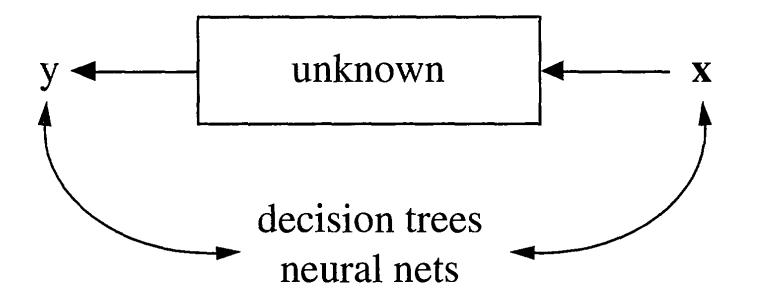
y nature x

Classical statistical modeling



"Assume data are generated by the following model..."

Machine Learning (Supervised)



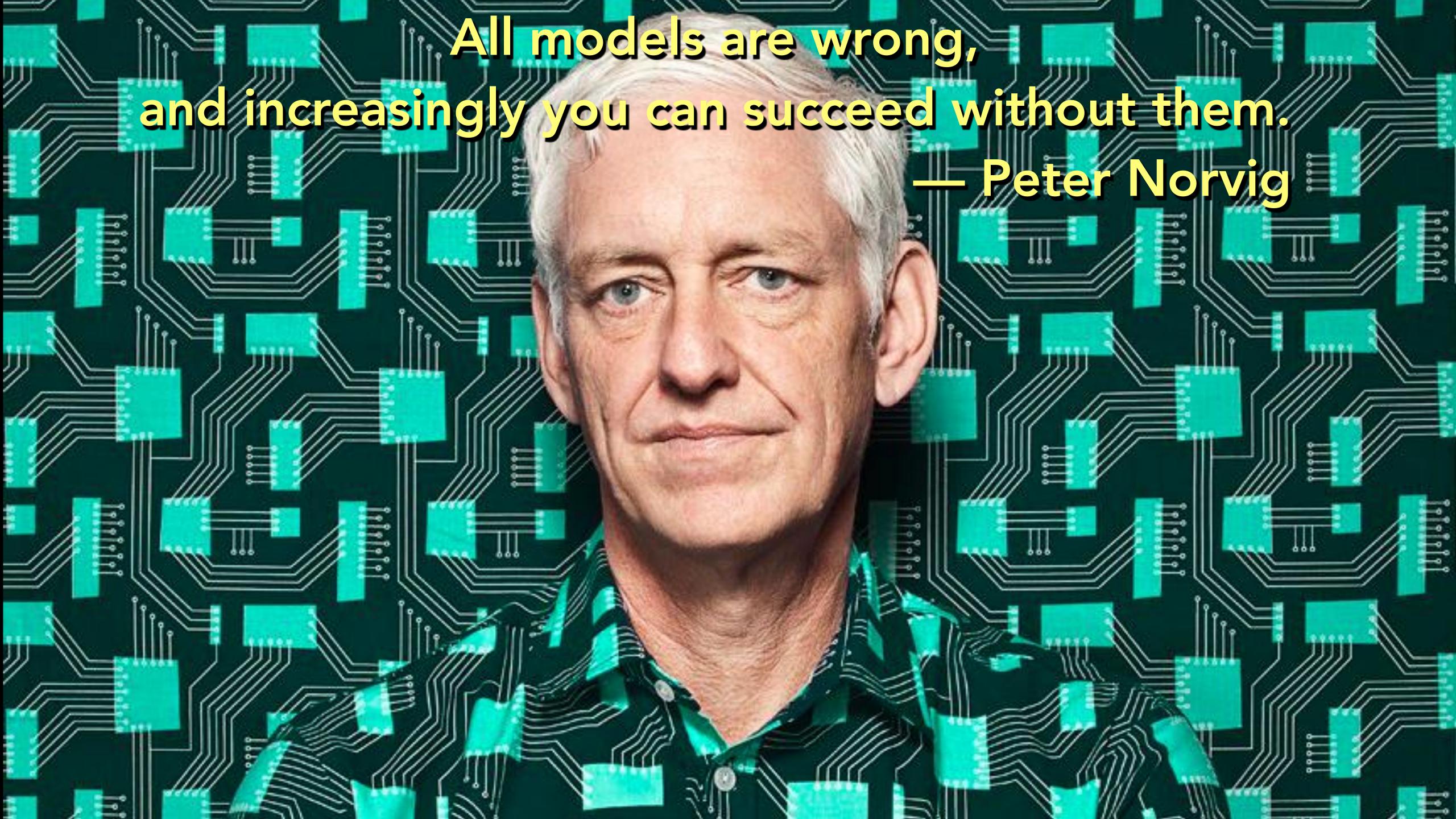
"Model-Free"
Black box
Non-parametric
Curve fitting

+ cross validation

7.1 A New Research Community

In the mid-1980s two powerful new algorithms for fitting data became available: neural nets and decision trees. A new research community using these tools sprang up. Their goal was predictive accuracy. The community consisted of young computer scientists, physicists and engineers plus a few aging statisticians. They began using the new tools in working on complex prediction problems where it was obvious that data models were not applicable: speech recognition, image recognition, nonlinear time series prediction, handwriting recognition, prediction in financial markets.

Their interests range over many fields that were once considered happy hunting grounds for statisticians and have turned out thousands of interesting research papers related to applications and methodology. A large majority of the papers analyze real data. The criterion for any model is what is the predictive accuracy. An idea of the range of research of this group can be got by looking at the Proceedings of the Neural Information Processing Systems Conference (their main yearly meeting) or at the Machine Learning Journal.



THE END OF THEORY: THE DATA DELUGE MAKES THE SCIENTIFIC METHOD OBSOLETE



A more nuanced view eScience & the 4th paradigm

Jim Gray and eScience in 2007



eScience -- A Transformed Scientific Method



Jim Gray, eScience Group, Microsoft Research

http://research.microsoft.com/~Gray

in collaboration with **Alex Szalay**

Dept. Physics & Astronomy Johns Hopkins University

http://www.sdss.jhu.edu/~szalay/



Jim Gray and eScience in 2007

Science Paradigms

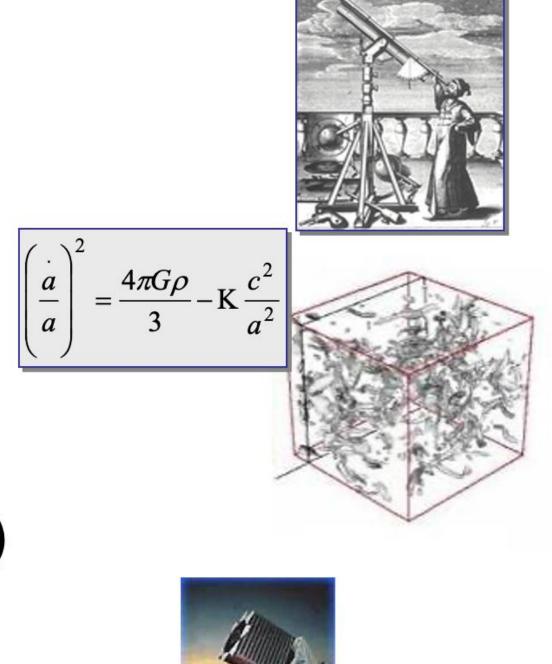
- Thousand years ago: science was empirical describing natural phenomena
- Last few hundred years:
 theoretical branch using models, generalizations
- Last few decades:

 a computational branch
 simulating complex phenomena
- Today:

data exploration (eScience)

unify theory, experiment, and simulation

- Data captured by instruments
 Or generated by simulator
- Processed by software
- Information/Knowledge stored in computer
- Scientist analyzes database / files using data management and statistics





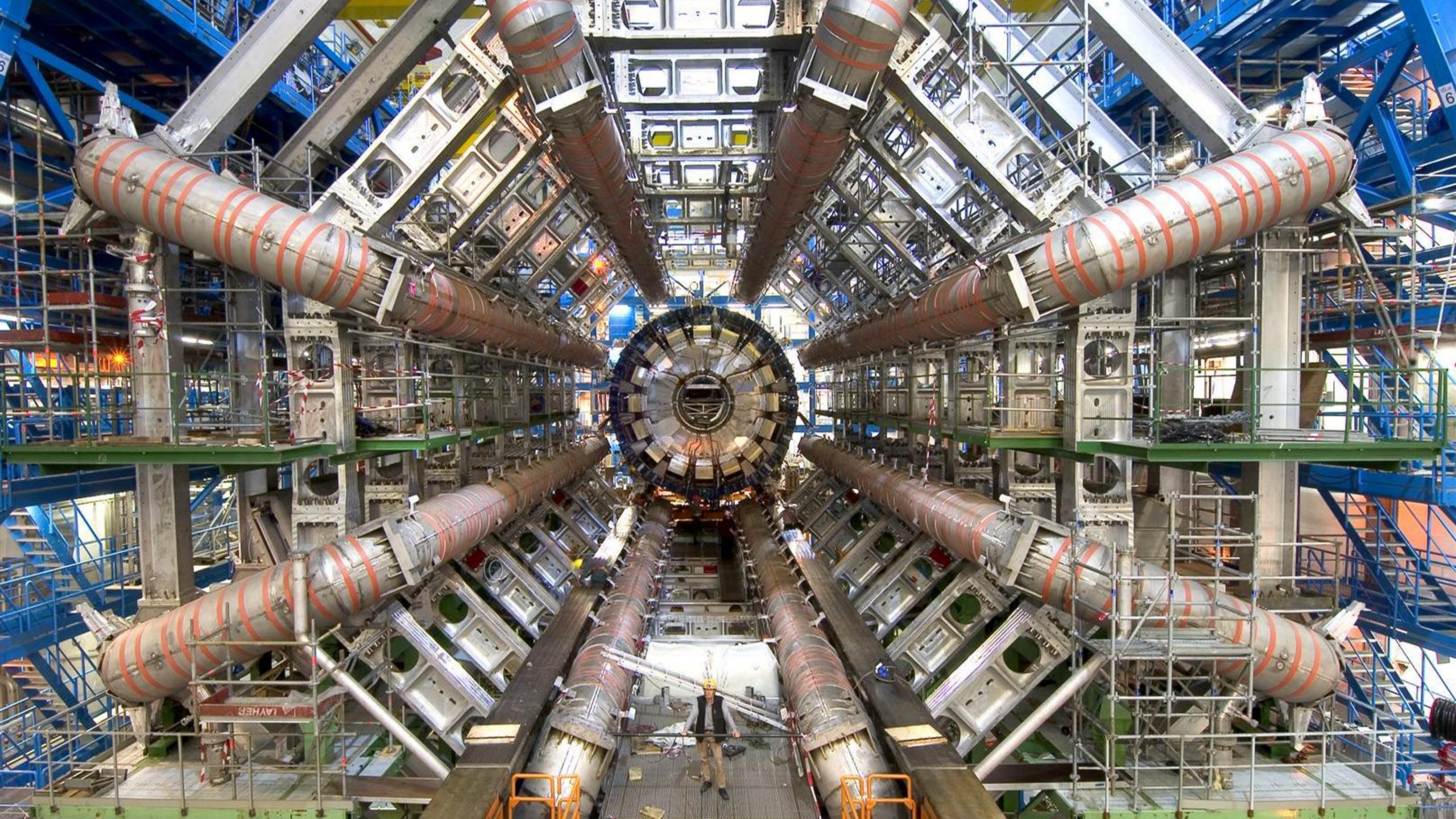


FOURTH
PARADIGM

DATA-INTENSIVE SCIENTIFIC DISCOVERY

Scientific Paradigms:

- 1. Empirical
- 2. Theoretical
- 3. Computational
- 4. Data-Driven / Data-Intensive Discovery



Expert feature engineering

Don't believe the media:

$$E \neq mc^2$$

What Einstein really said:

$$E^{2} = (mc^{2})^{2} + (|\vec{p}|c)^{2}$$

Every physics student knows energy and momentum are conserved

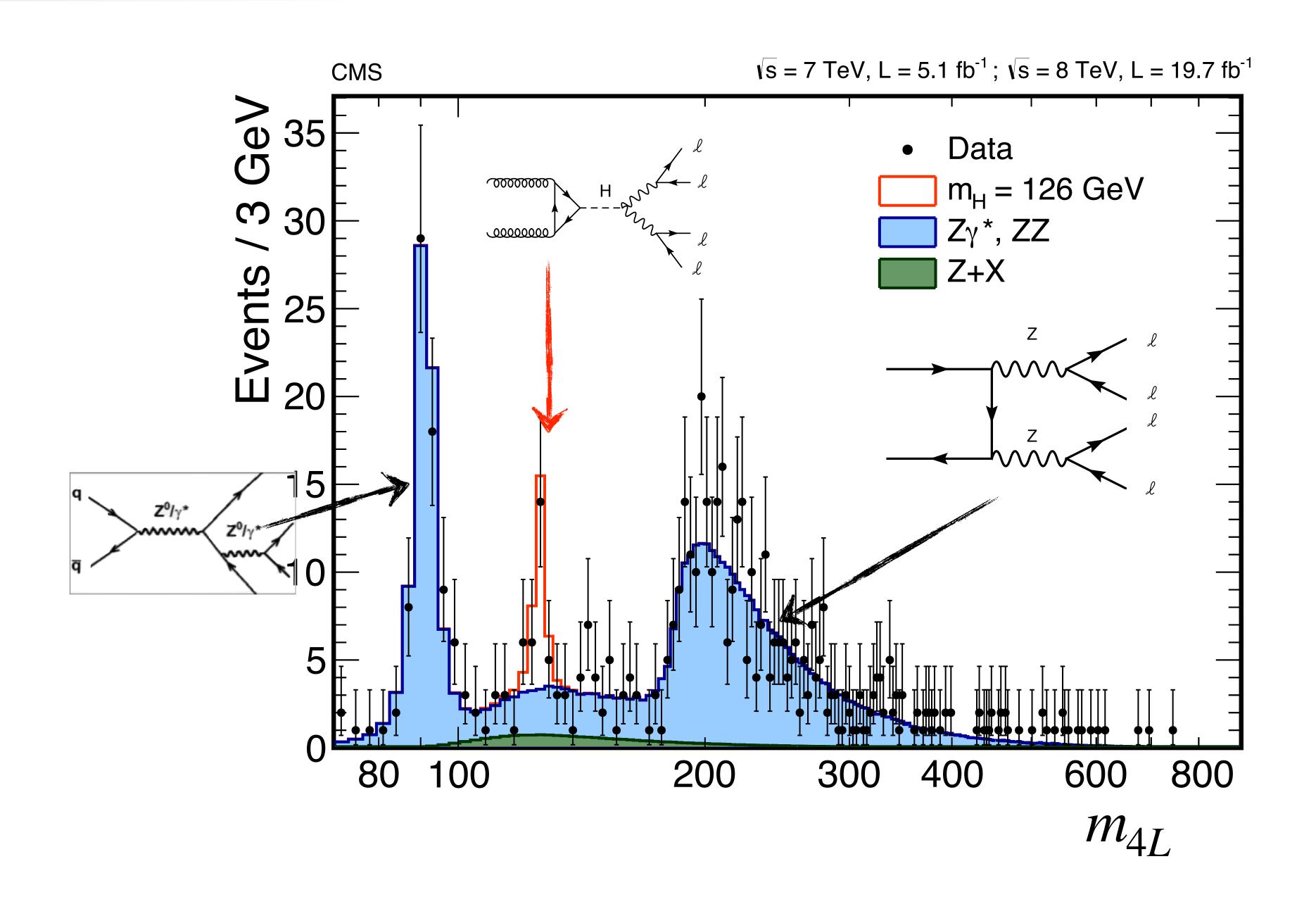
$$E_{\text{Higgs}} = E_{\text{before}} = E_{\text{after}} = \sum_{i} E_{i}$$

 $\vec{p}_{\text{Higgs}} = \vec{p}_{\text{before}} = \vec{p}_{\text{after}} = \sum_{i} \vec{p}_{i}$

Thus, for our hypothesis, we expect a peak in this feature / observable:

$$m_{4L} = \sqrt{E_{\text{after}}^2/c^4 - |\vec{p}_{\text{after}}|^2/c^2}$$

A signal emerges

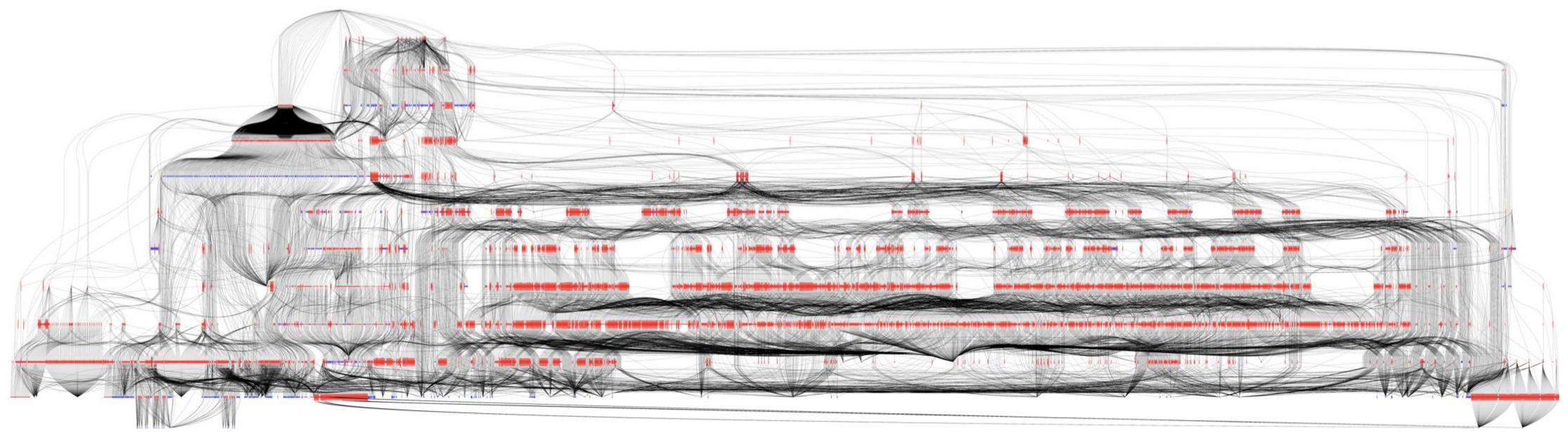


Developing the statistical approaches for the LHC

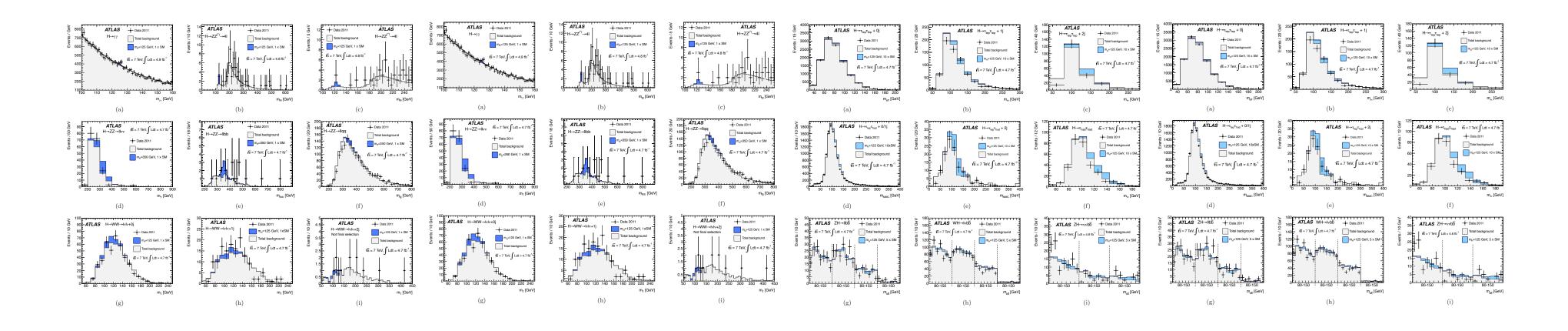
From 2003-2010, I was mainly busy developing a **formalism** and a software **implementation** that allowed us **to combine** (statistically) **evidence** from **multiple datasets**

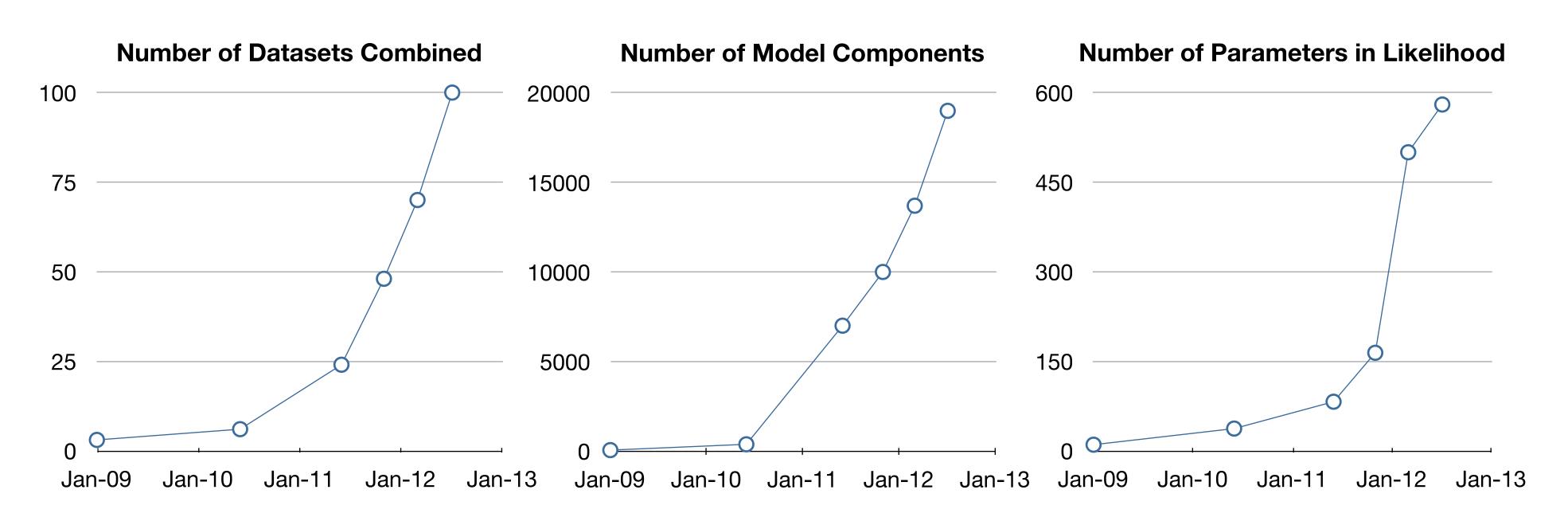
 Mainly frequentist approach that could cope with complicated models with many nuisance parameters

$$\mathbf{f}_{\text{tot}}(\mathcal{D}_{\text{sim}}, \mathcal{G}|\boldsymbol{\alpha}) = \prod_{c \in \text{channels}} \left[\text{Pois}(n_c|\nu_c(\boldsymbol{\alpha})) \prod_{e=1}^{n_c} f_c(x_{ce}|\boldsymbol{\alpha}) \right] \cdot \prod_{p \in \mathbb{S}} f_p(a_p|\alpha_p)$$



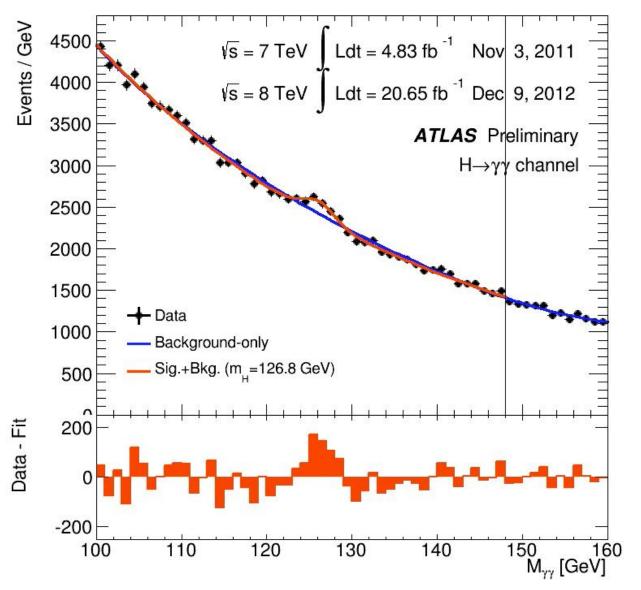
Collaborative Statistical Modeling

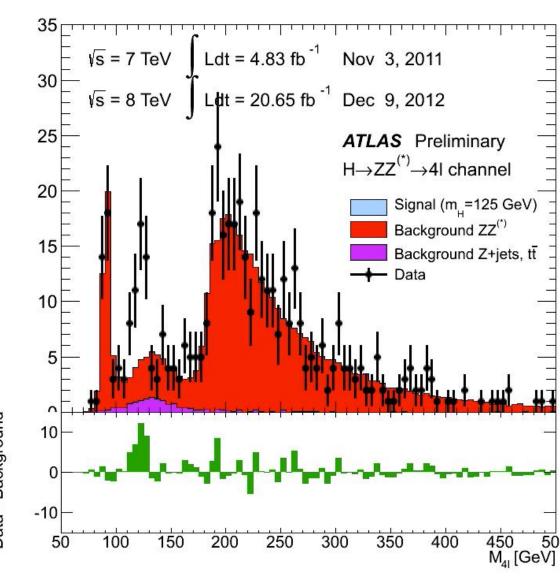




July 2012

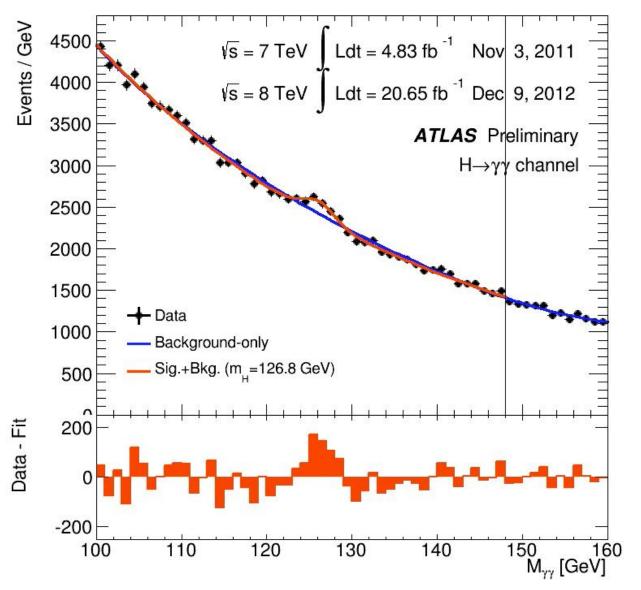


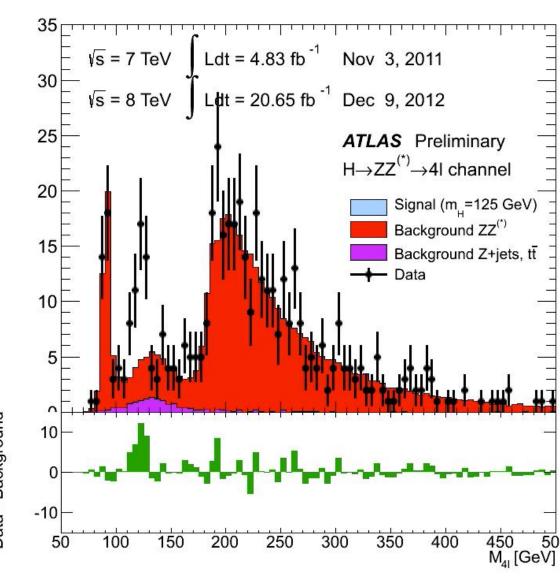




July 2012

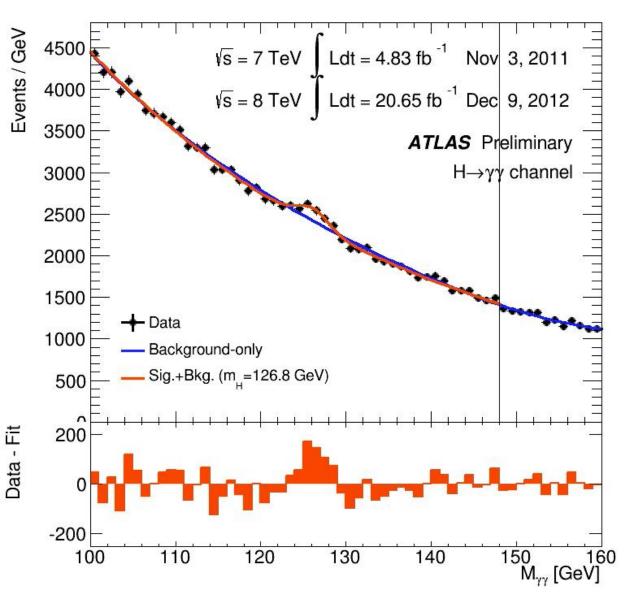


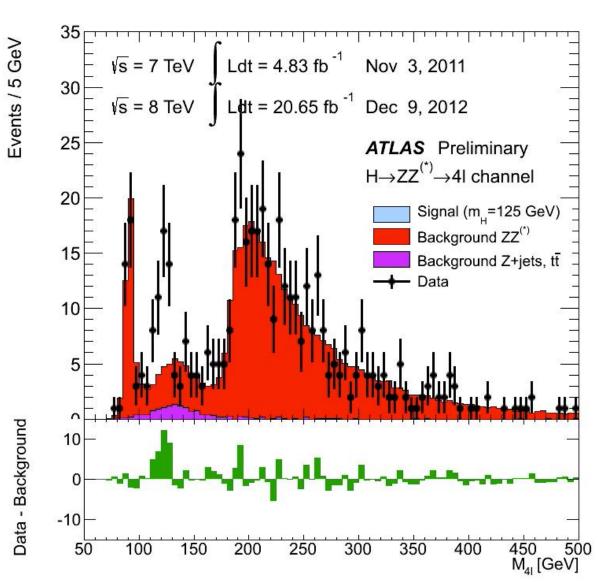




July 2012







ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya SutskeverUniversity of Toronto

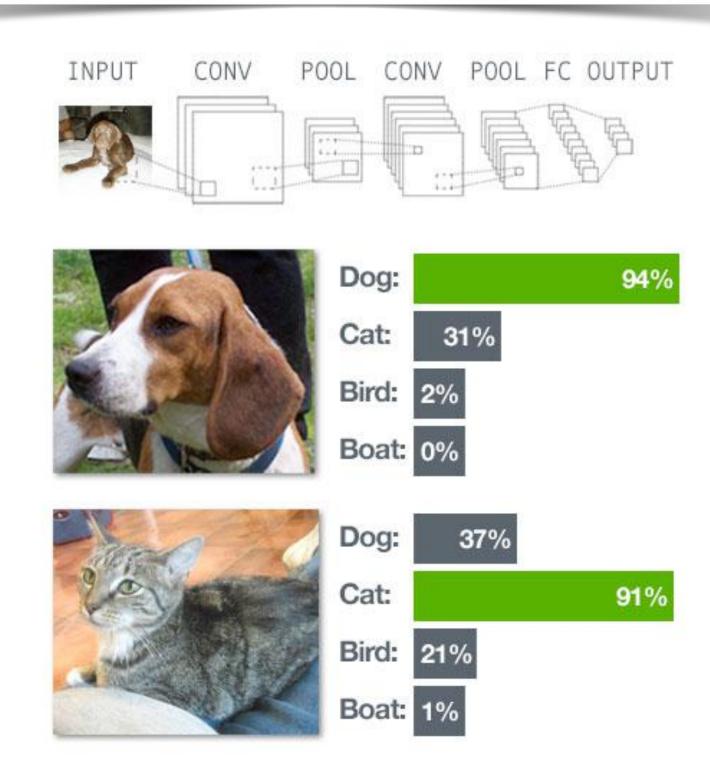
ilya@cs.utoronto.ca

Geoffrey E. HintonUniversity of Toronto

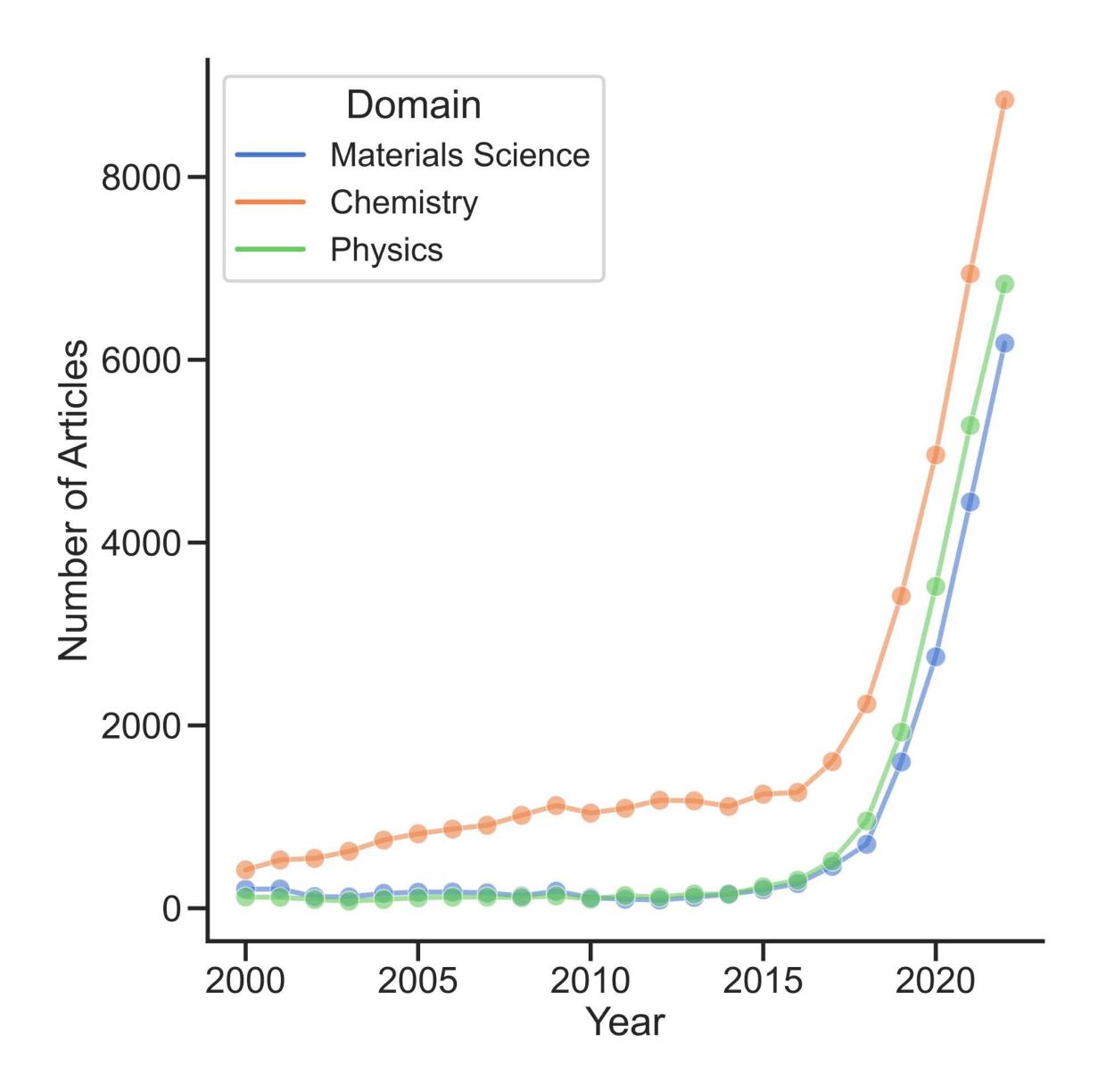
hinton@cs.utoronto.ca

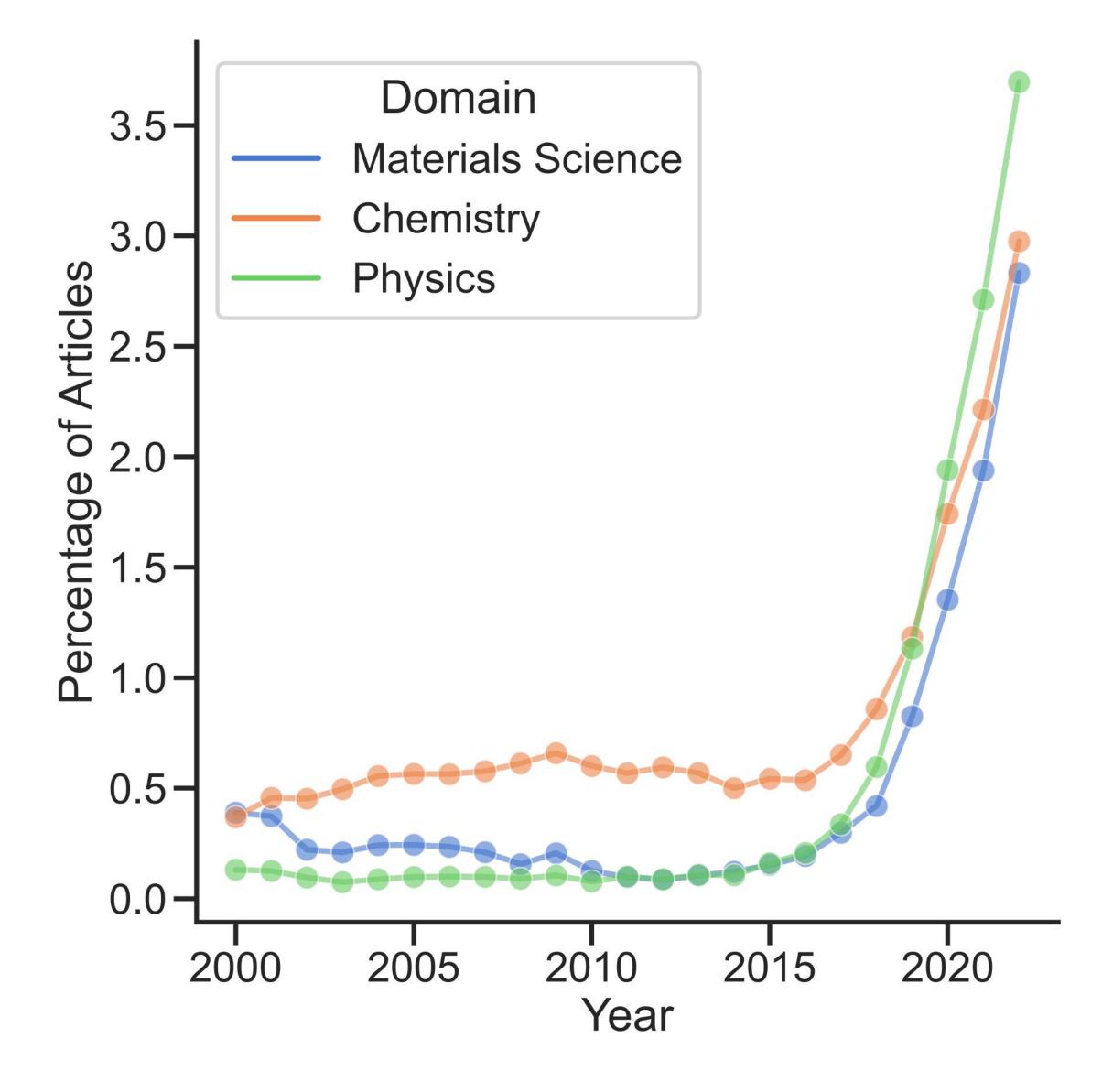
Abstract

We trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation. To reduce overfitting in the fully-connected layers we employed a recently-developed regularization method called "dropout" that proved to be very effective. We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry.



ML Publications in Science

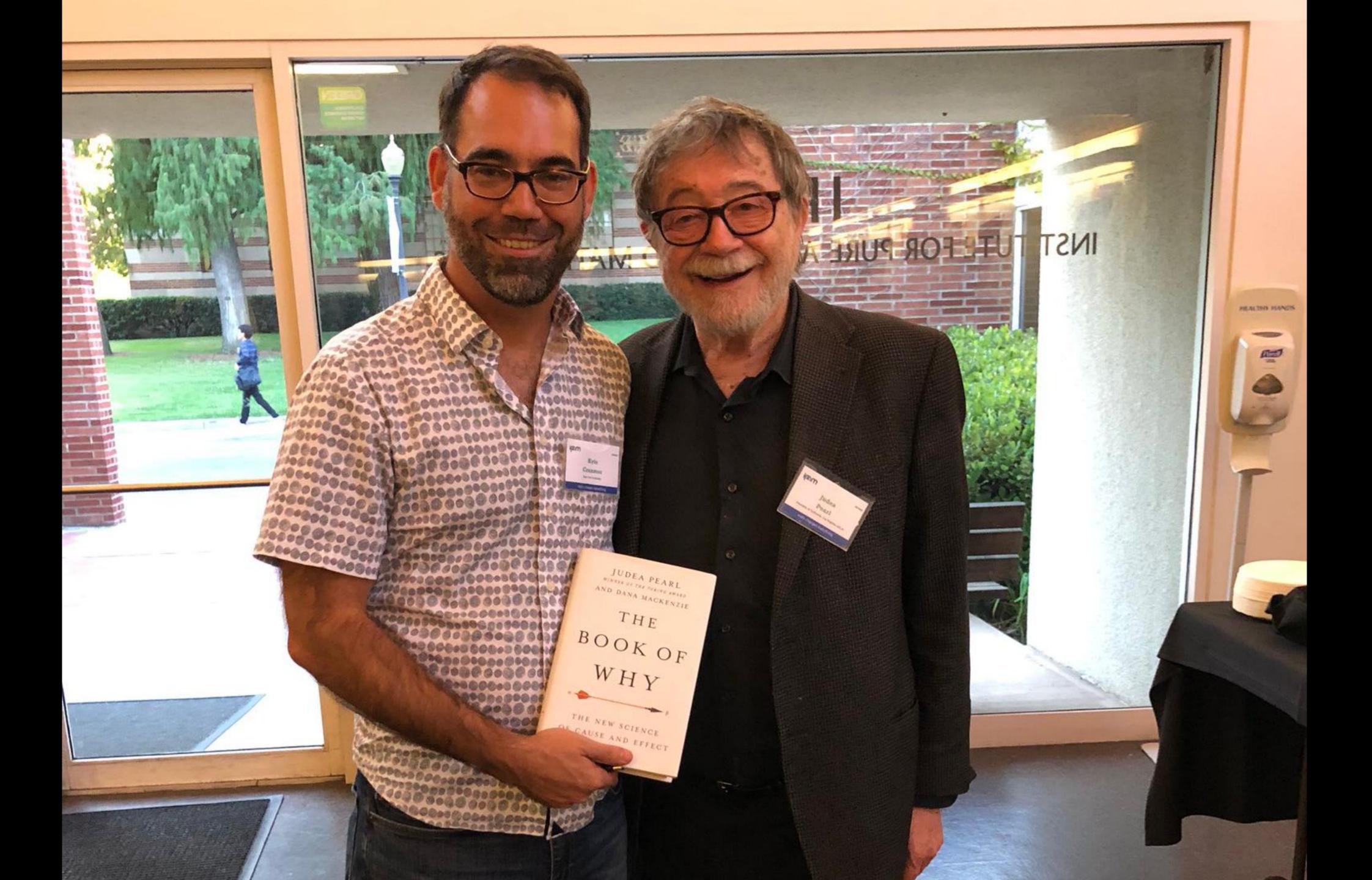


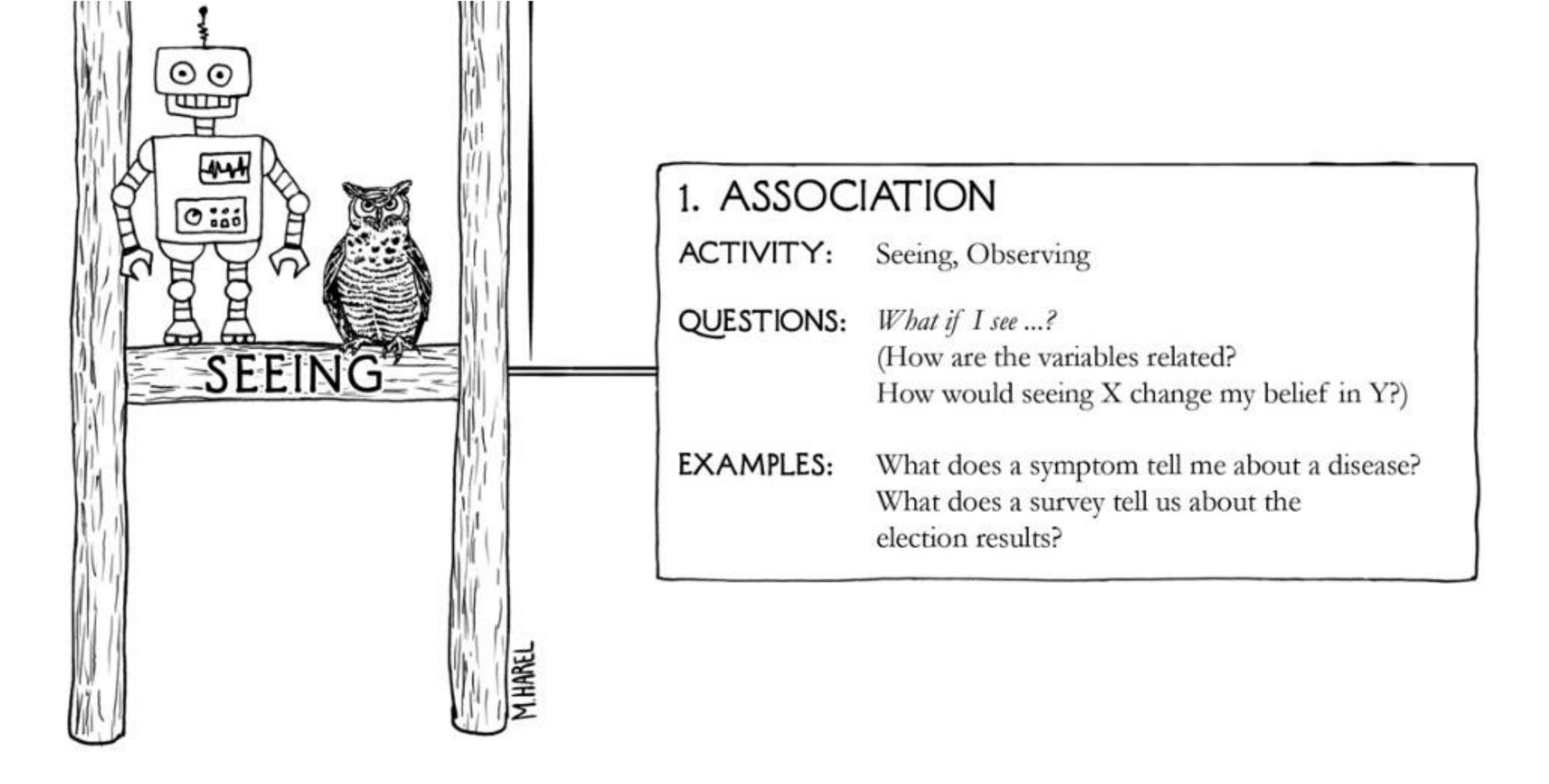


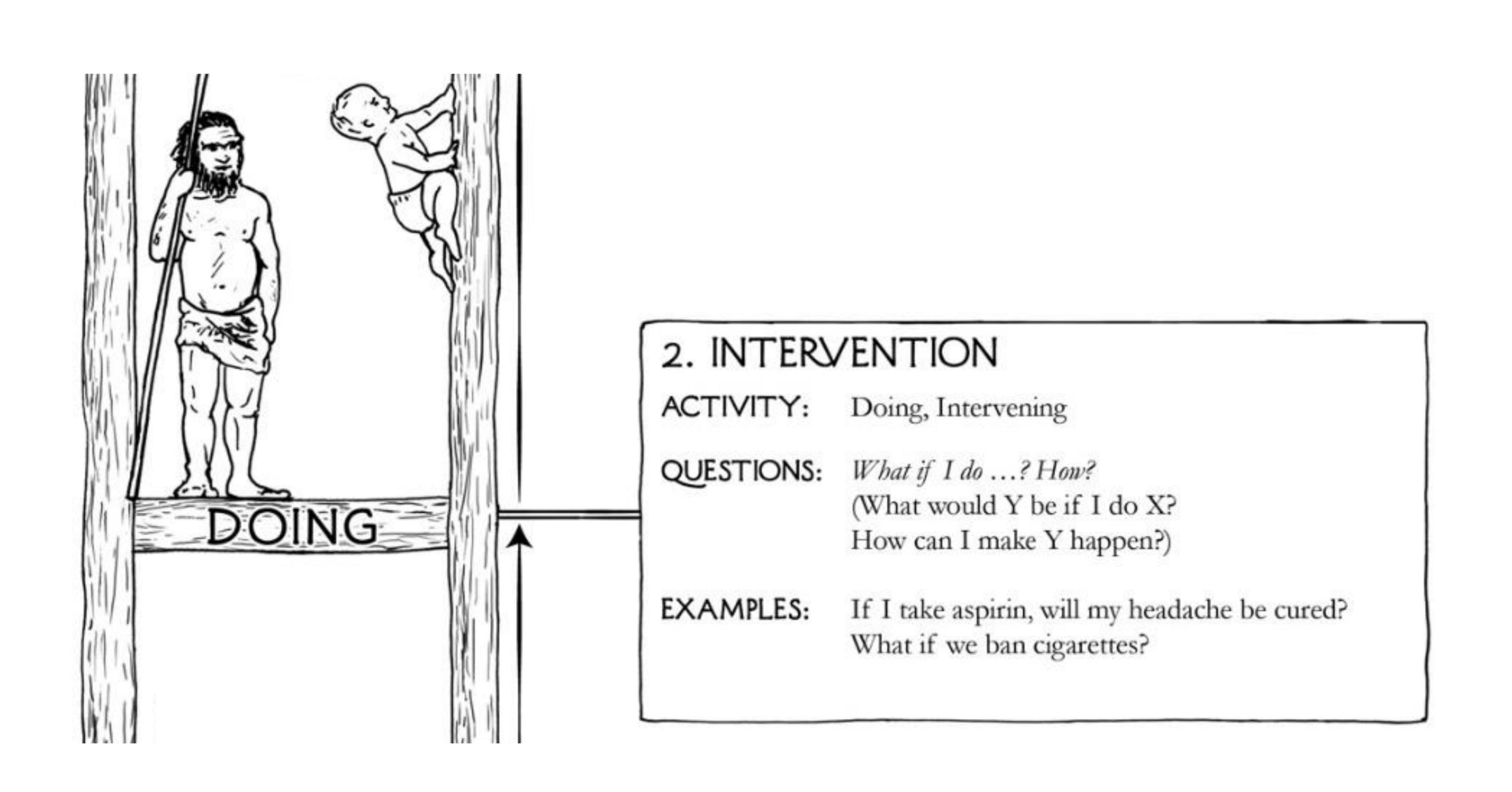
Data + ML isn't enough!

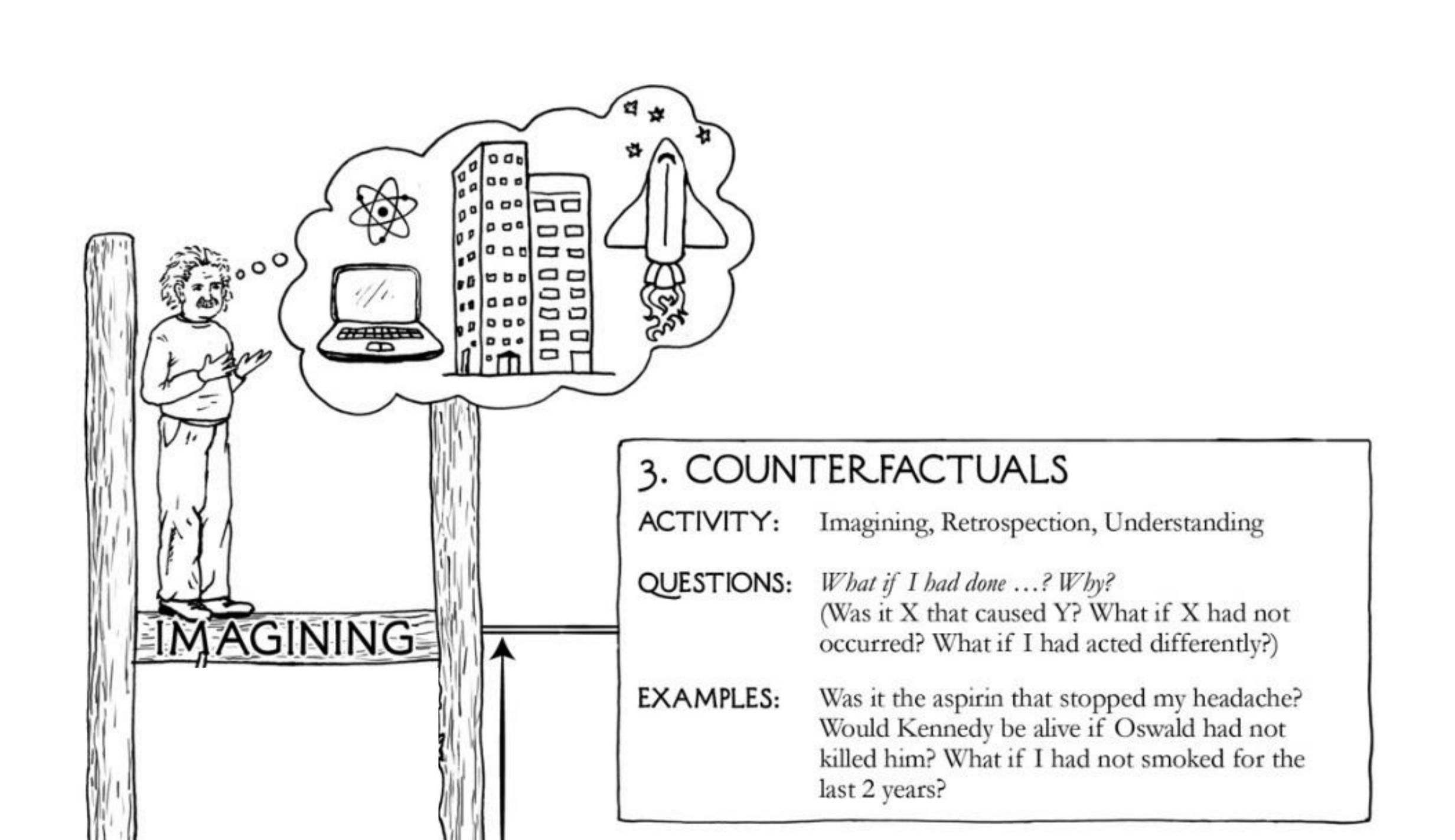
Why isn't data + ML enough?

Causation vs. Correlation









3 scenarios with the same correlation

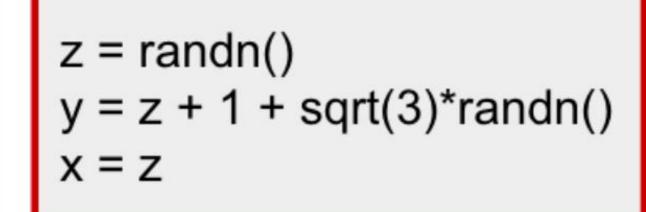


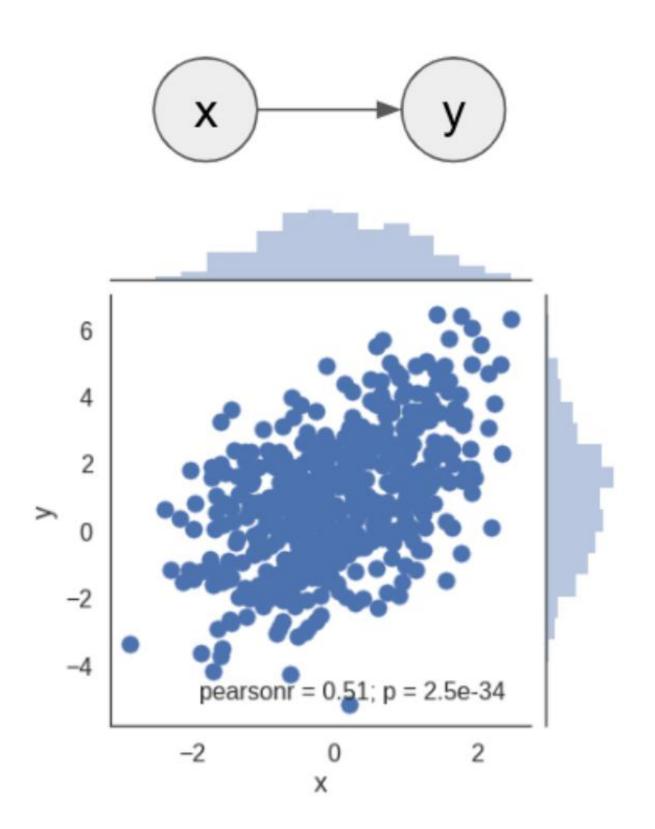
$$x = randn()$$

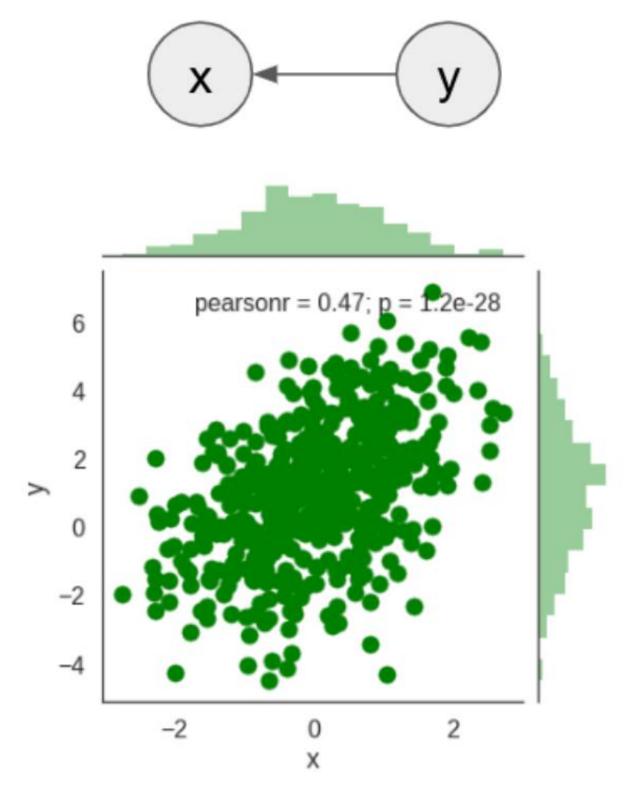
 $y = x + 1 + sqrt(3)*randn()$

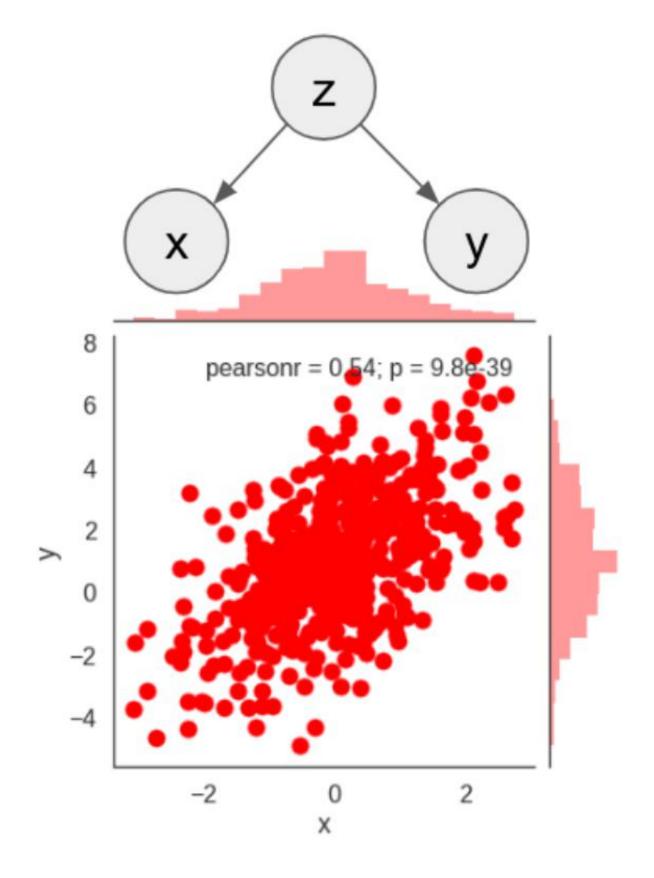
$$y = 1 + 2*randn()$$

 $x = (y-1)/4 + sqrt(3)*randn()/2$



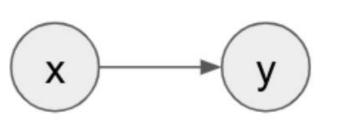


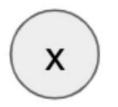


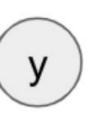


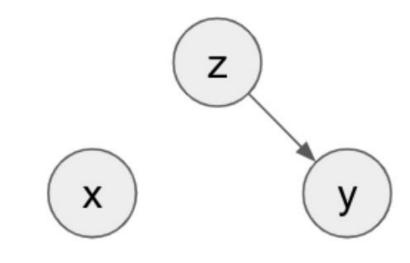
...but different response to intervention







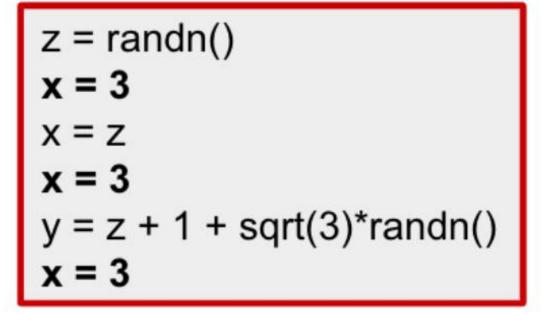


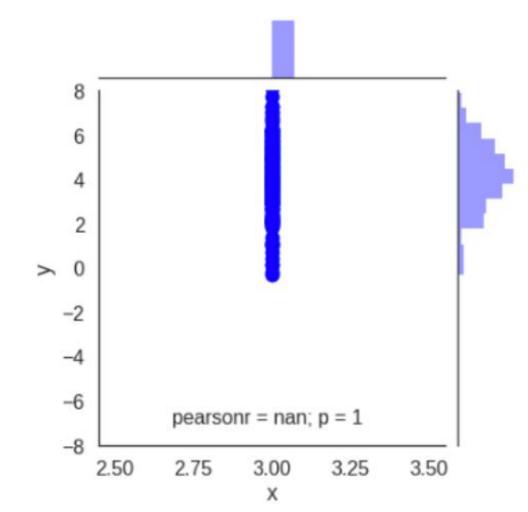


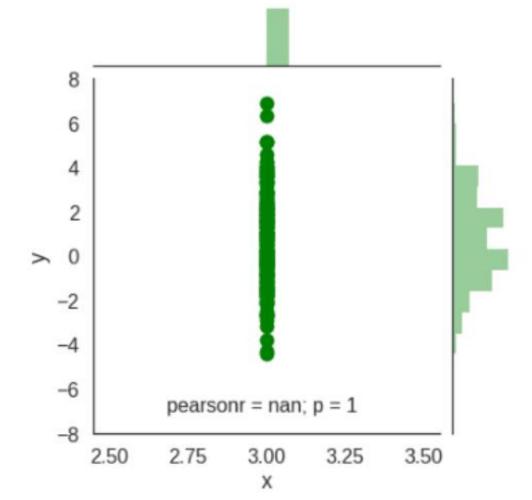
$$P(y|do(X)) = p(y|x)$$

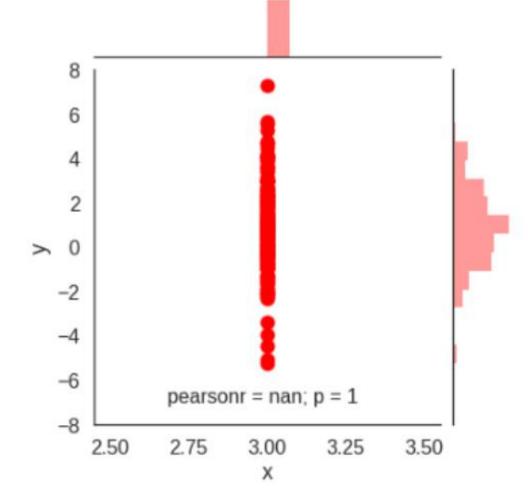
$$P(y|do(X)) = p(y)$$

$$P(y|do(X)) = p(y)$$

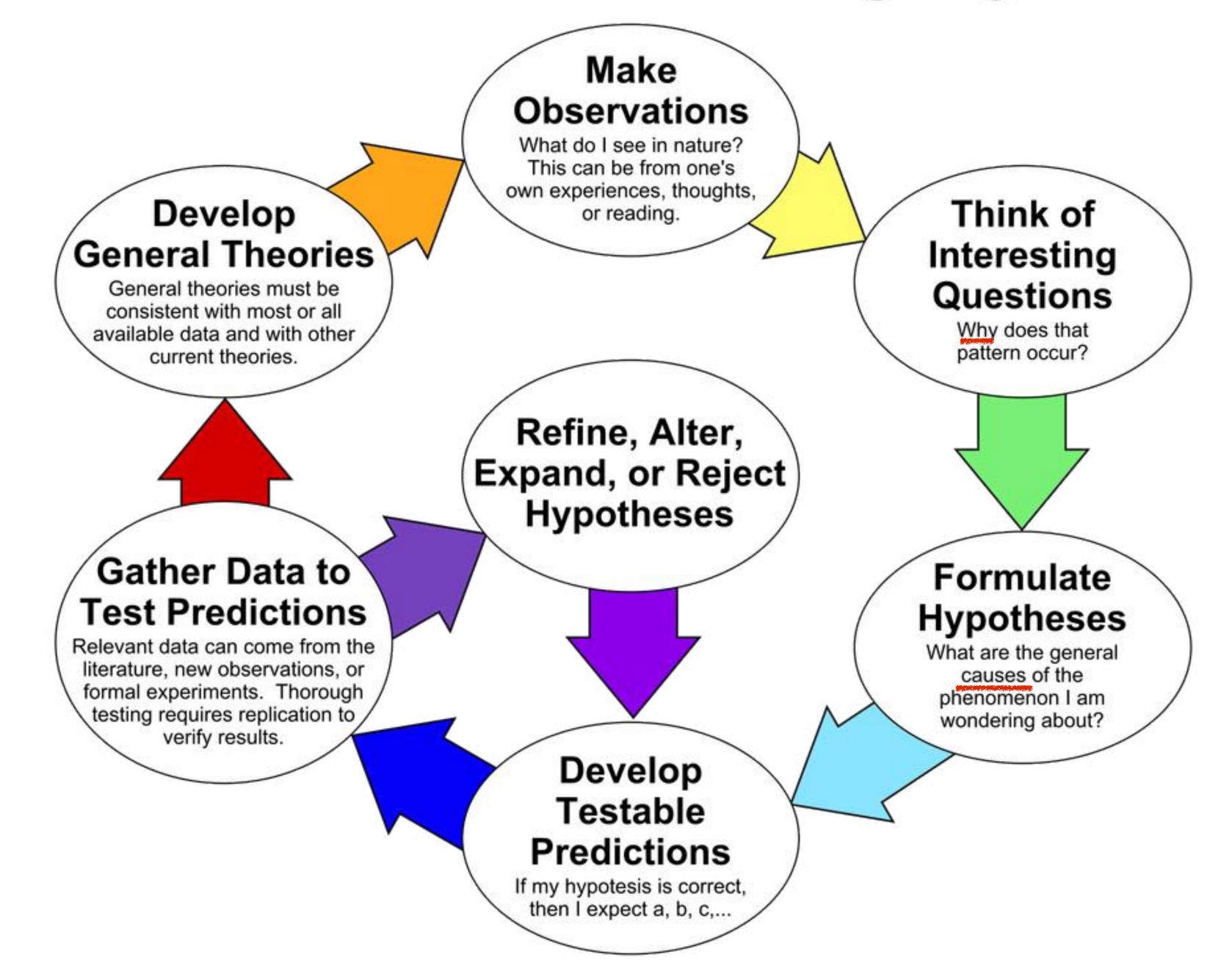








The Scientific Method as an Ongoing Process



Scientific Understanding

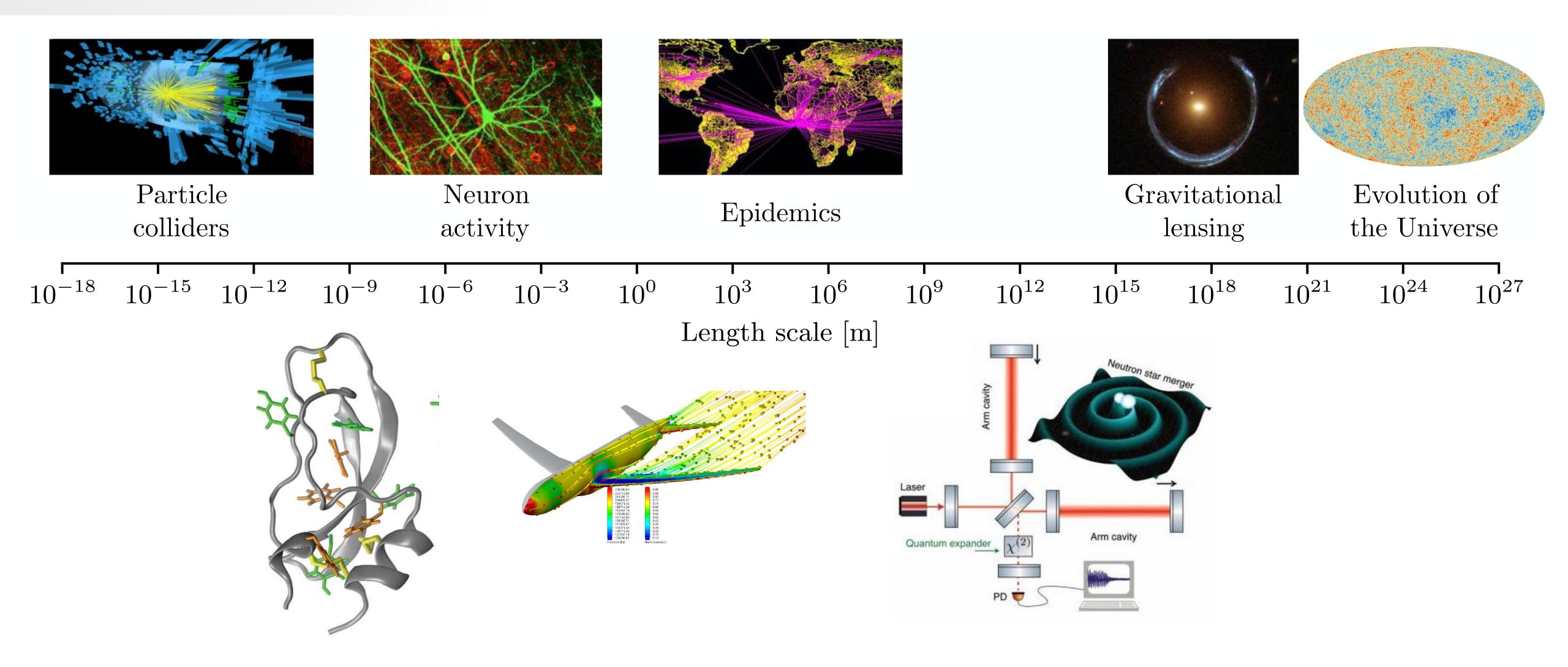
Breiman's view contrasts with this current view

On scientific understanding with artificial intelligence

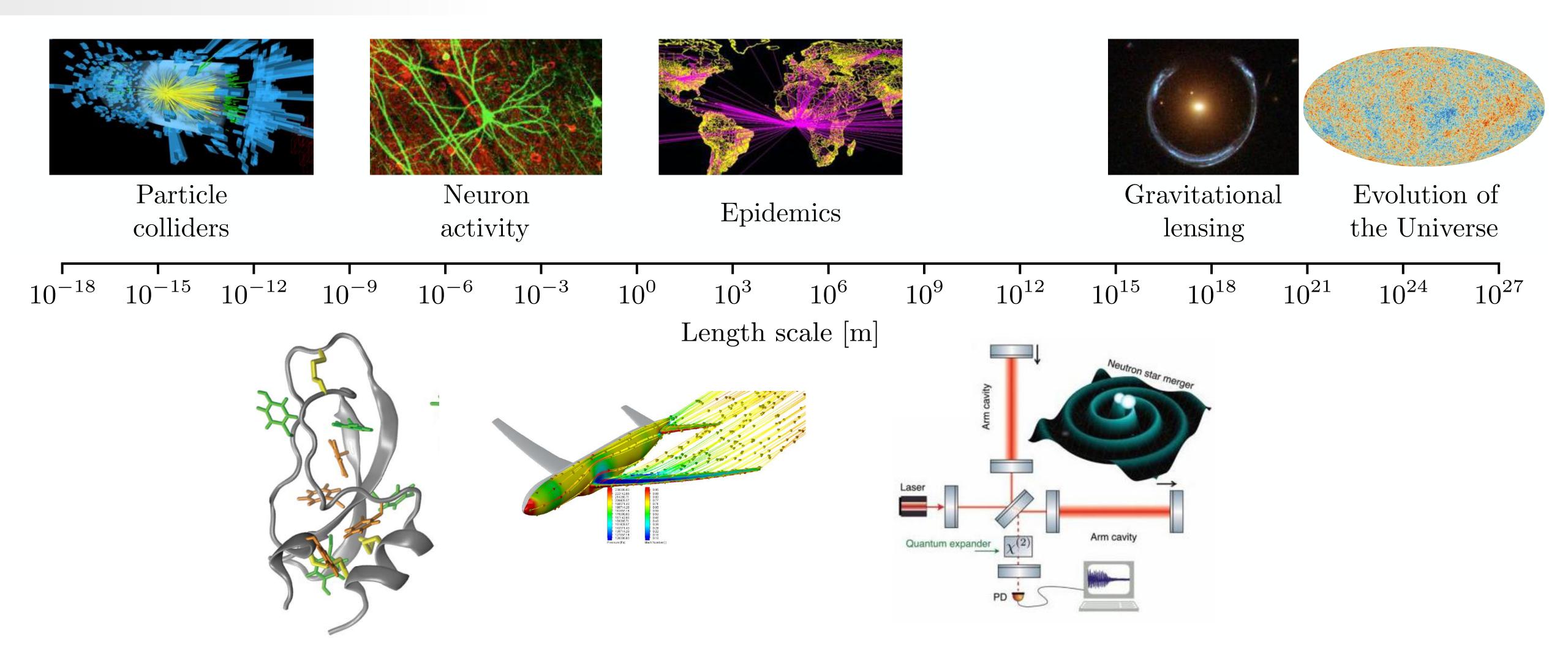
Mario Krenn, ^{1,2,3,4,*} Robert Pollice, ^{2,3} Si Yue Guo, ² Matteo Aldeghi, ^{2,3,4} Alba Cervera-Lierta, ^{2,3} Pascal Friederich, ^{2,3,5} Gabriel dos Passos Gomes, ^{2,3} Florian Häse, ^{2,3,4,6} Adrian Jinich, ⁷ Akshat Kumar Nigam, ^{2,3} Zhenpeng Yao, ^{2,8,9,10} and Alán Aspuru-Guzik ^{2,3,4,11,†}

Imagine an oracle that <u>correctly predicts</u> the outcome of every particle physics experiment, the products of every chemical reaction, or the function of every protein. Such an oracle would revolutionize science and technology as we know them. However, as scientists, we would not be satisfied with the oracle itself. We want more. We want to comprehend how the oracle conceived these predictions. This feat, denoted as scientific understanding, has frequently been recognized as the essential aim of science. Now, the ever-growing power of computers and artificial intelligence poses one ultimate question: How can advanced artificial systems contribute to scientific understanding or achieve it autonomously?

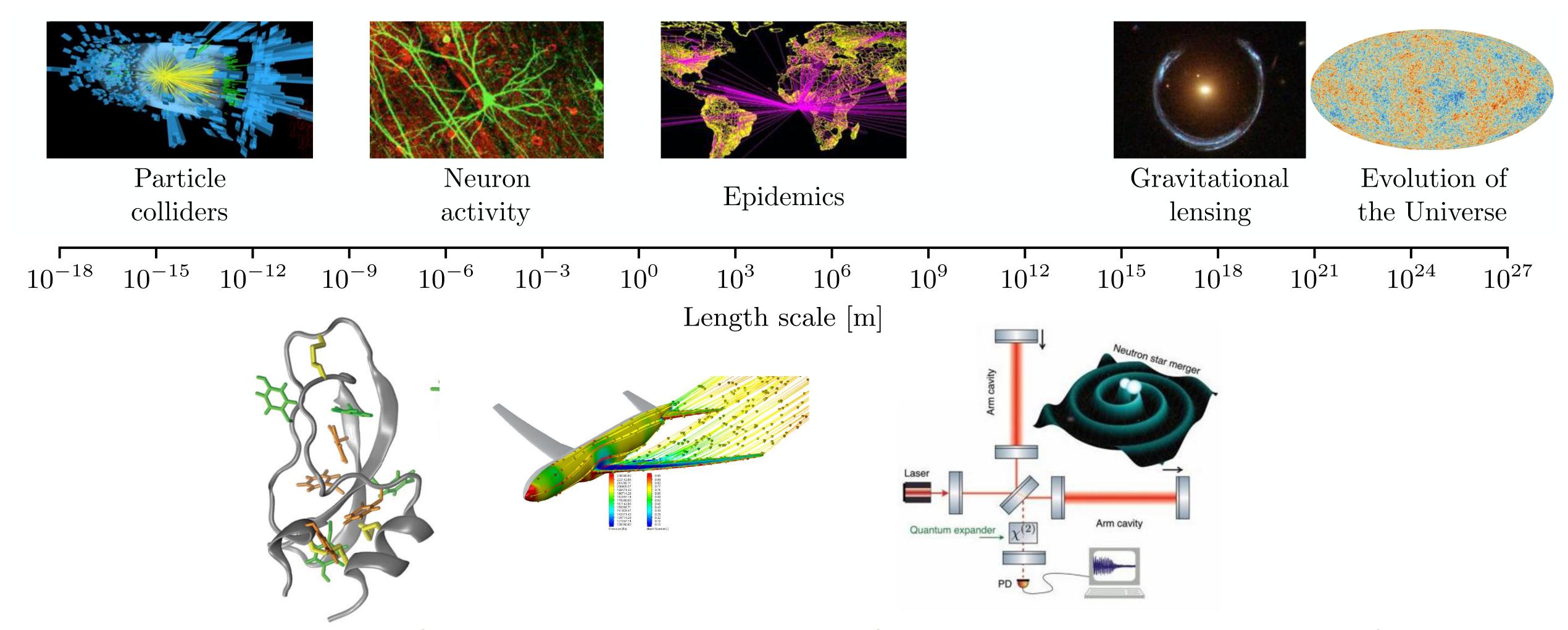
A fifth paradigm?



The forefront of scientific knowledge is often encapsulated in simulators

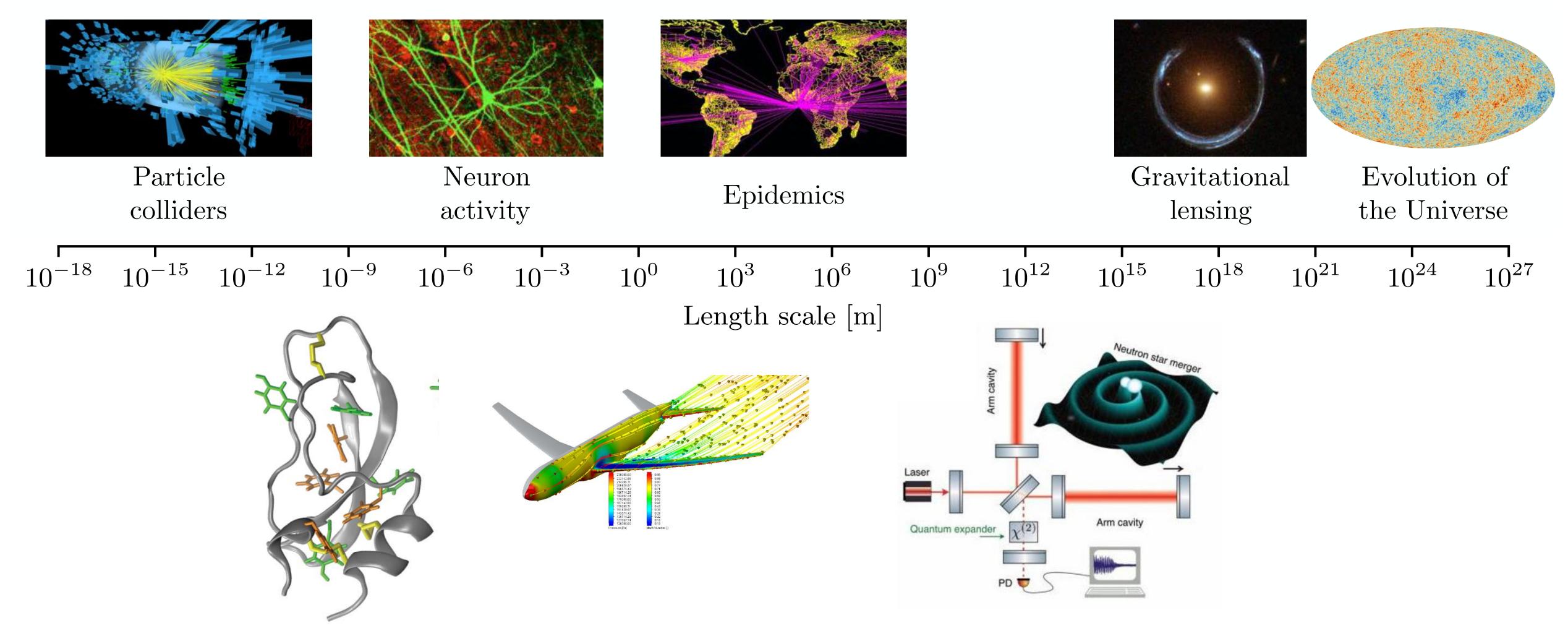


Simulators are causal, generative models of the data generating process

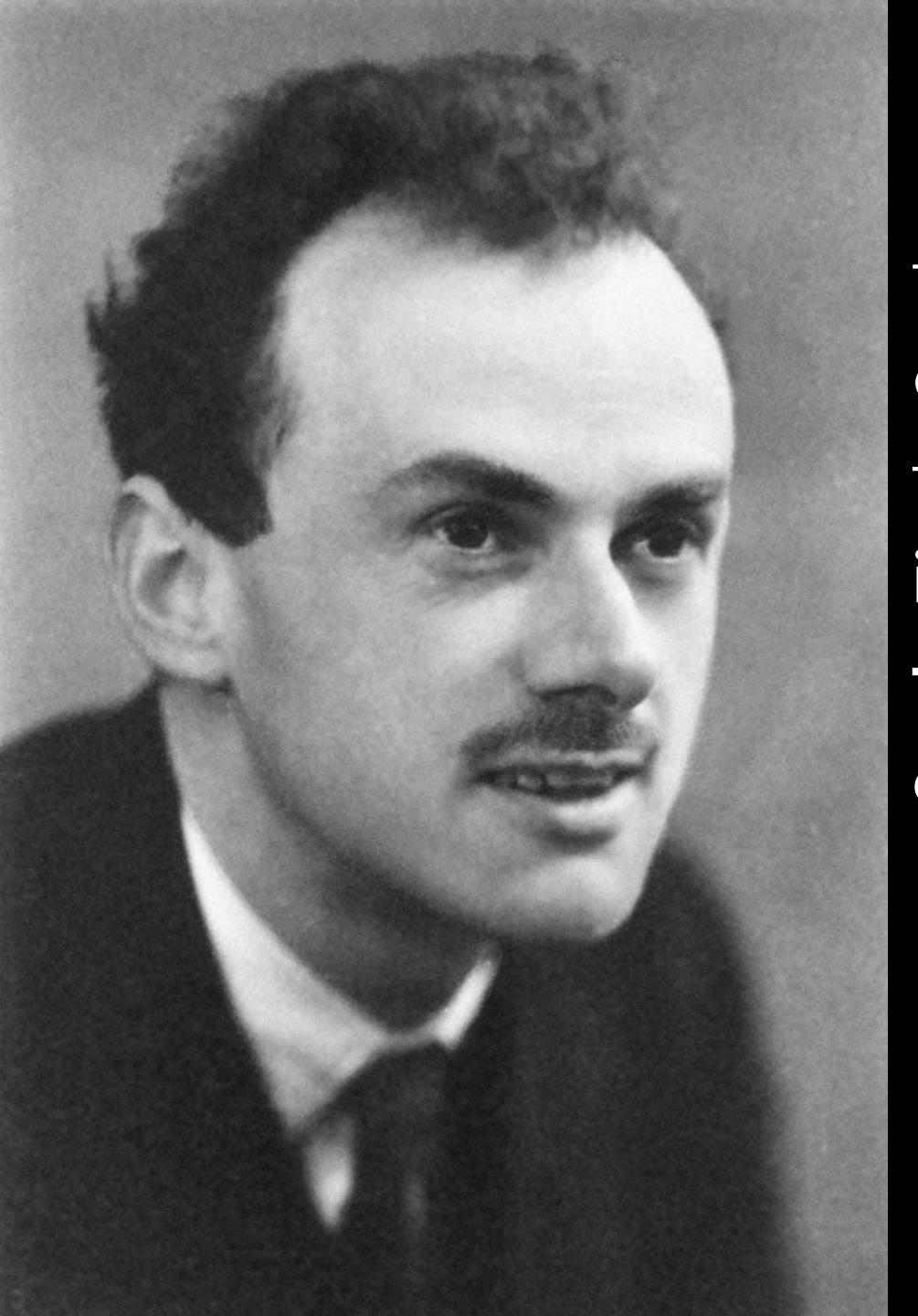


The expressiveness of programming languages facilitates the development of complex, high-fidelity simulations, and the power of modern computing provides the ability to generate synthetic data from them.

[Cranmer, Brehmer, Louppe PNAS (2020), arXiv:1911.01429] 37



Unfortunately, these simulators are poorly suited for many downstream tasks, e.g. statistical inference, experimental design, decision making, ...



"The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble."

-PAUL DIRAC

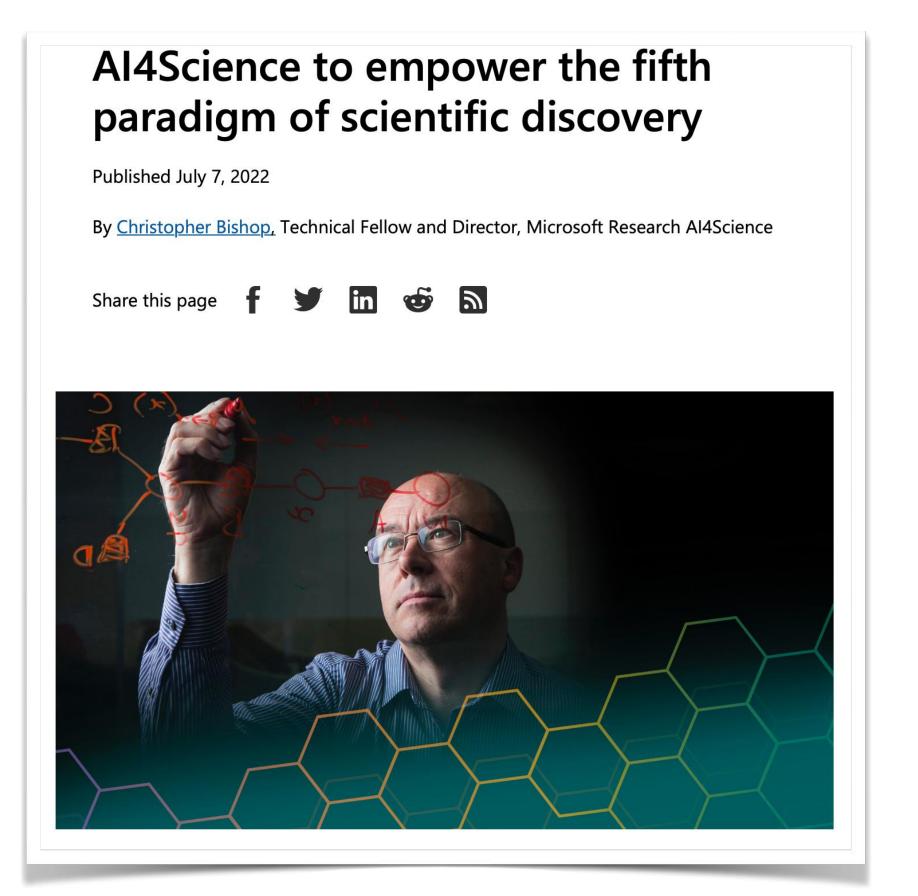


FOURTH
PARADIGM

DATA-INTENSIVE SCIENTIFIC DISCOVERY

Scientific Paradigms:

- 1. Empirical
- 2. Theoretical
- 3. Computational
- 4. Data-Driven / Data-Intensive Discovery
- 5. AI/ML + Simulation + Data



"New directions in science are launched by new tools much more often than by new concepts. The effect of a concept-driven revolution is to explain old things in new ways. The effect of a tool-driven revolution is to discover new things that have to be explained."

- FREEMAN DYSON



REALITY







Cow

Spherical Cow

ReLU Cow

In fusion energy



Accelerating fusion science through learned plasma control

February 16, 2022



Successfully controlling the nuclear fusion plasma in a tokamak with deep reinforcement learning

In molecular dynamics

Noé et al., Science **365**, 1001 (2019)

6 September 2019

RESEARCH

RESEARCH ARTICLE SUMMARY

MACHINE LEARNING

Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning

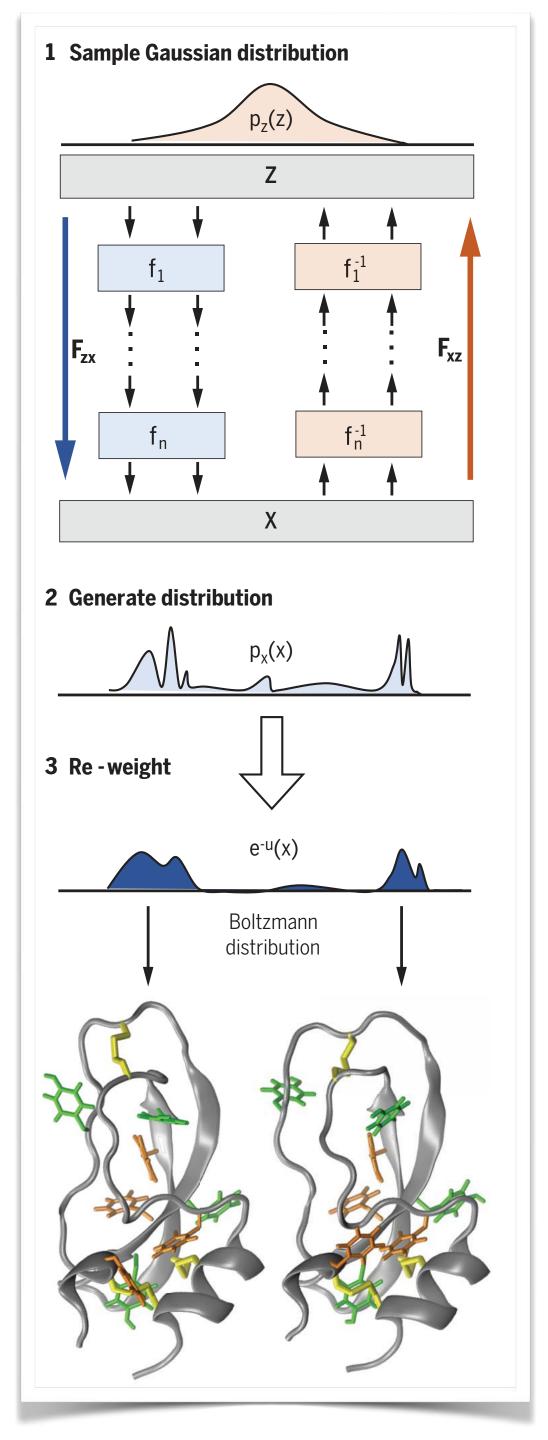
Frank Noé*†, Simon Olsson*, Jonas Köhler*, Hao Wu

The main approach is thus to start with one configuration, e.g., the folded protein state, and make tiny changes to it over time, e.g., by using Markov-chain Monte Carlo or molecular dynamics (MD). However, these simulations get trapped in metastable (long-lived) states: For example, sampling a single folding or unfolding event with atomistic MD may take a year on a supercomputer.



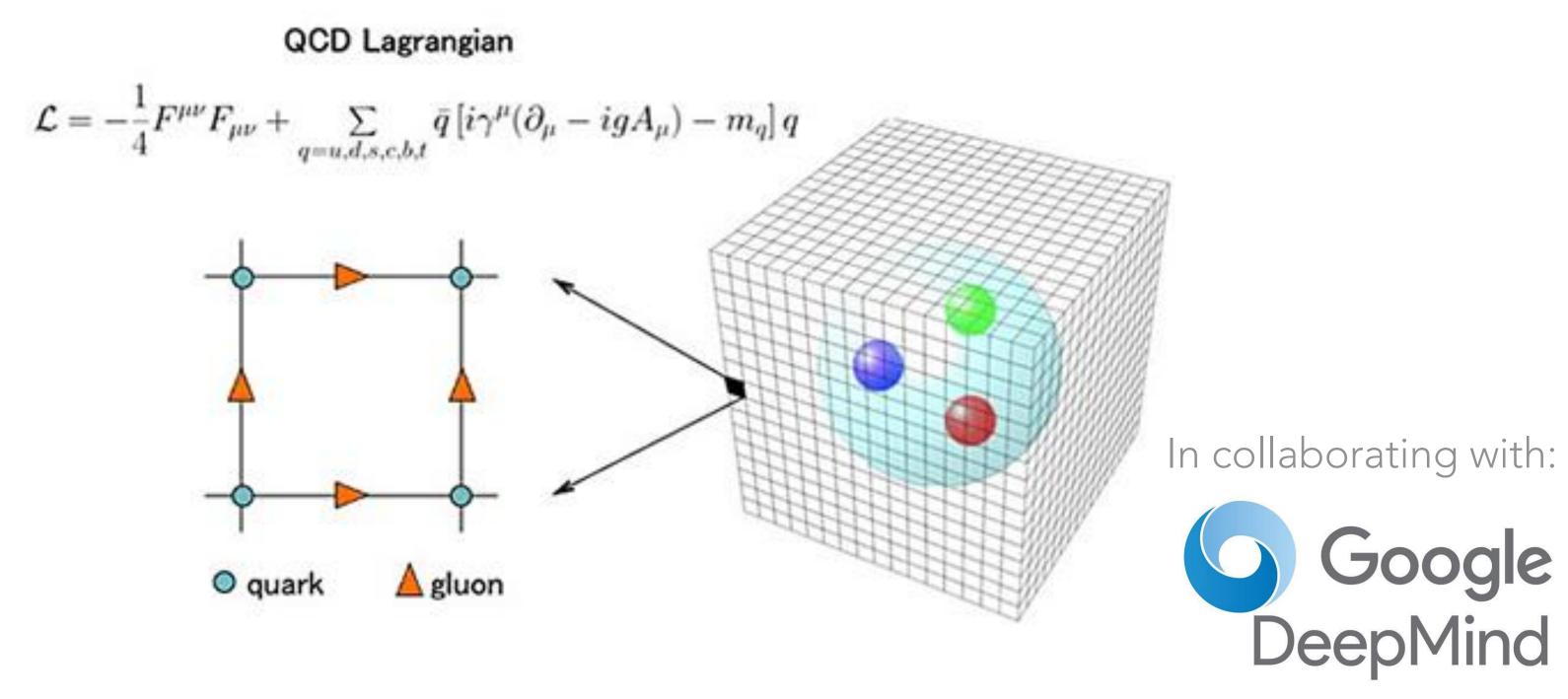


Boltzmann generators overcome sampling problems between long-lived states.



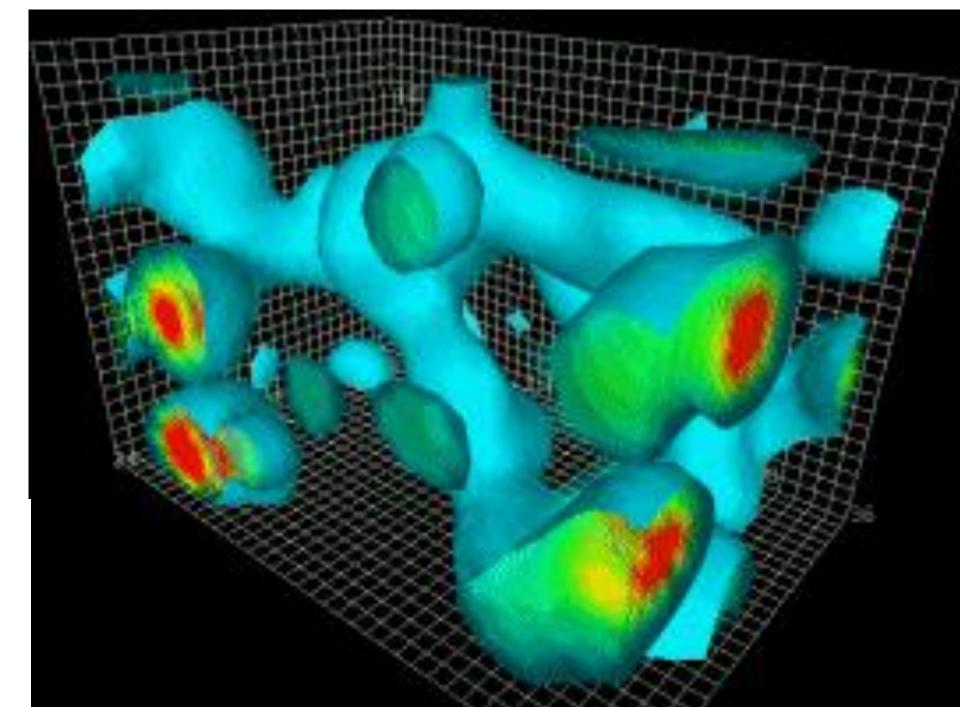
In theoretical nuclear physics & statistical physics

Extraordinarily computational intensive predictions for lattice quantum field theory



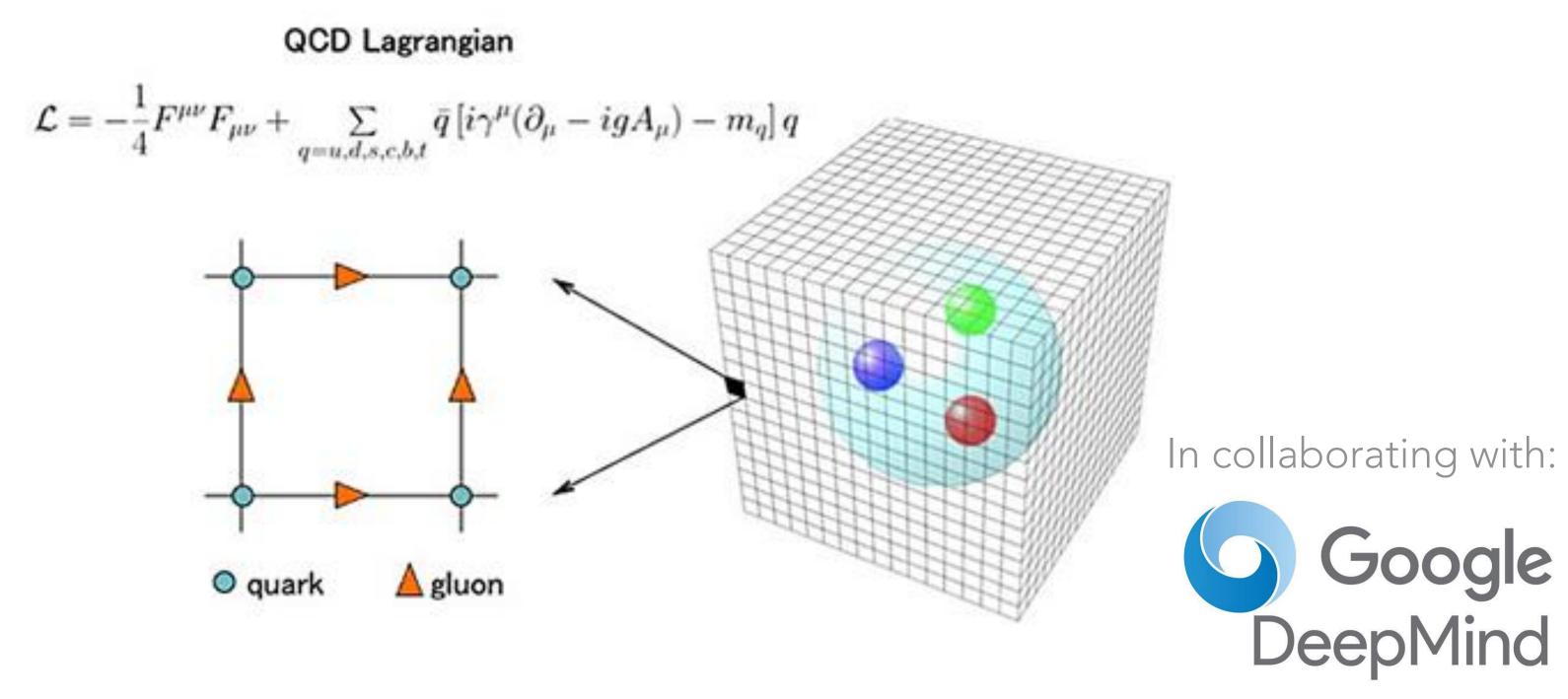


Nature Rev.Phys. 5 (2023) 9, 526-535; arXiv: 2305.02402; 2211.07541; : 2208.03832; 2207.08945; 2202.11712; 2107.00734; 2106.05934; 2008.05456; 2003.06413; 2002.02428; 1904.12072



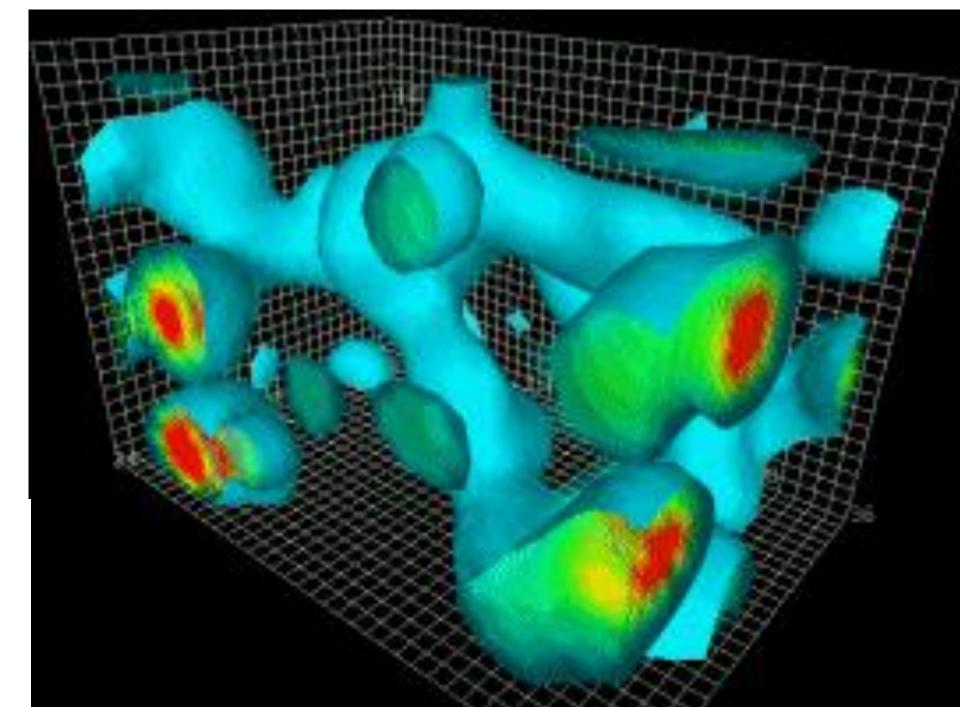
In theoretical nuclear physics & statistical physics

Extraordinarily computational intensive predictions for lattice quantum field theory



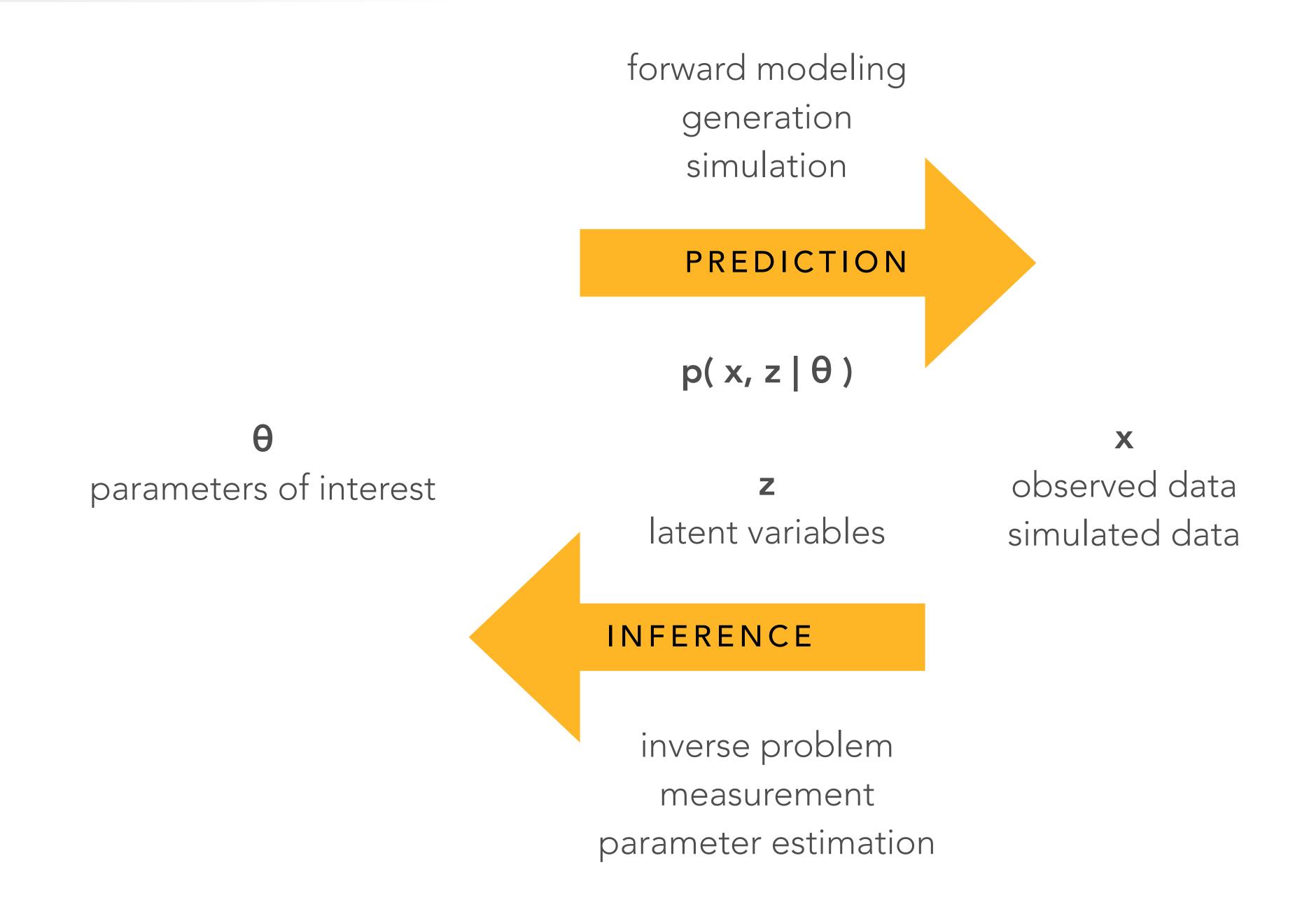


Nature Rev.Phys. 5 (2023) 9, 526-535; arXiv: 2305.02402; 2211.07541; : 2208.03832; 2207.08945; 2202.11712; 2107.00734; 2106.05934; 2008.05456; 2003.06413; 2002.02428; 1904.12072

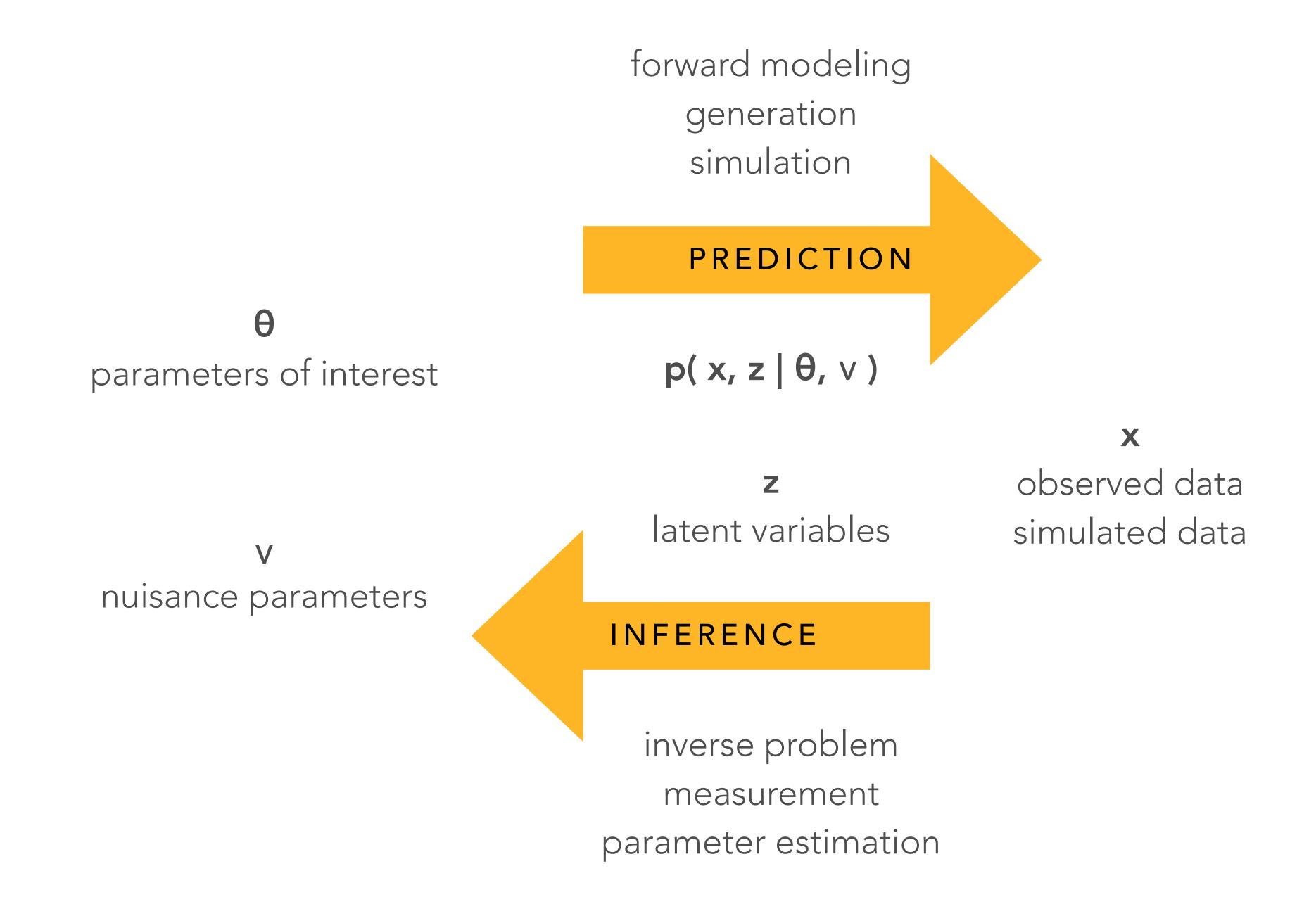


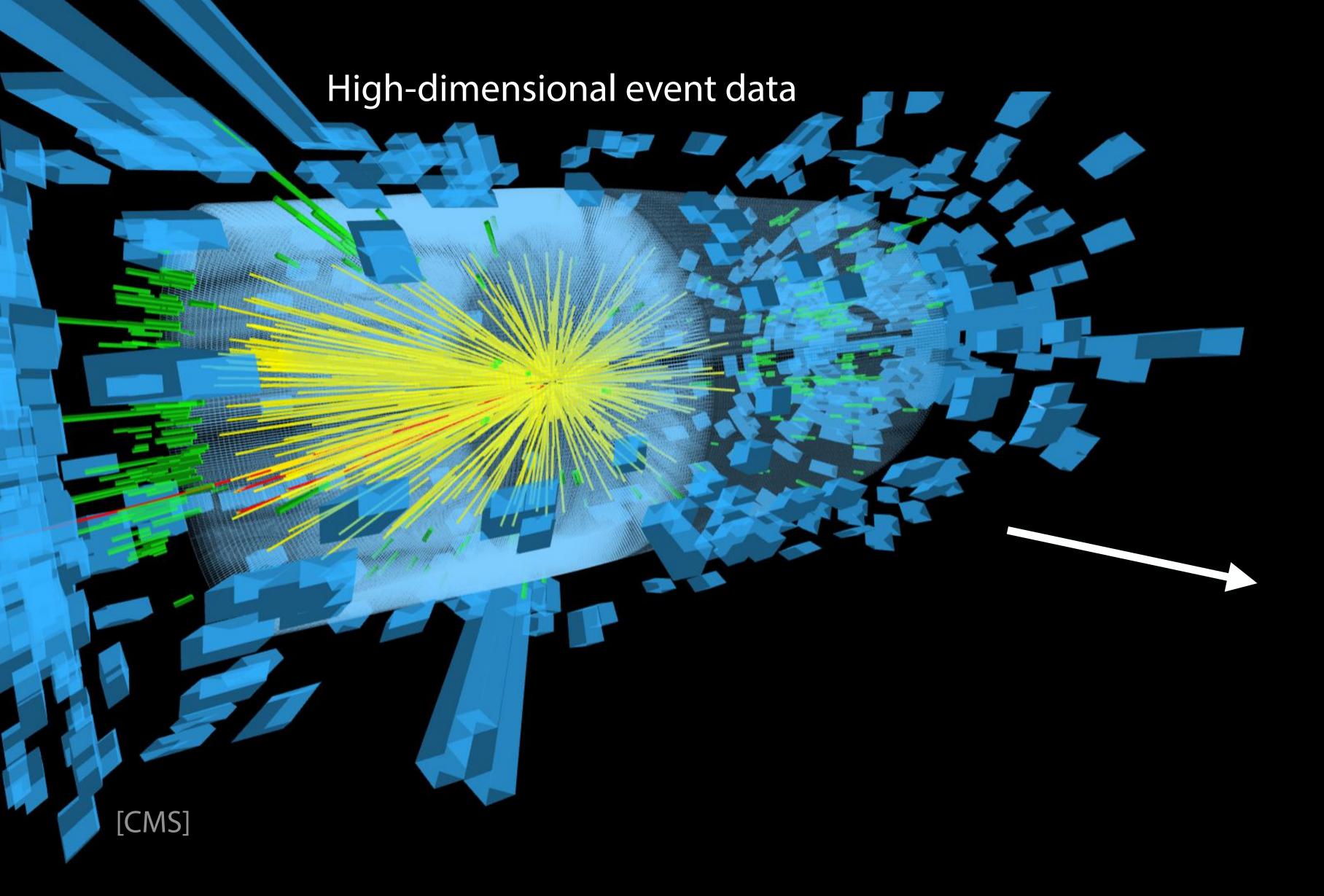
Simulation-Based Inference

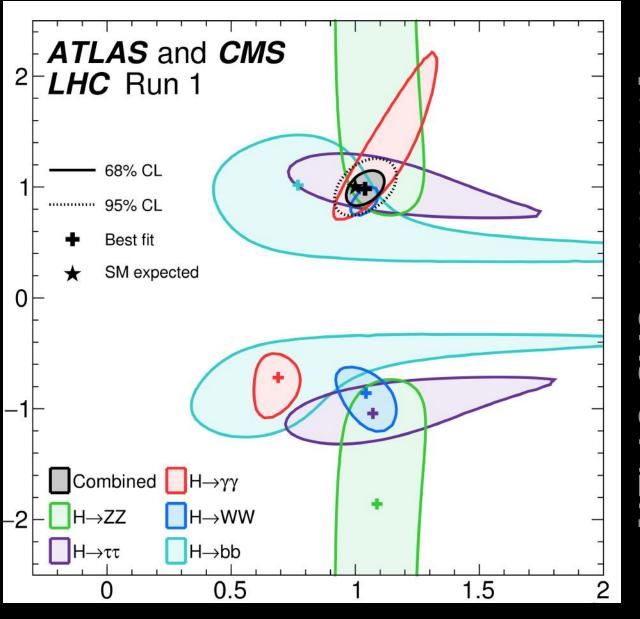
Statistical Framing



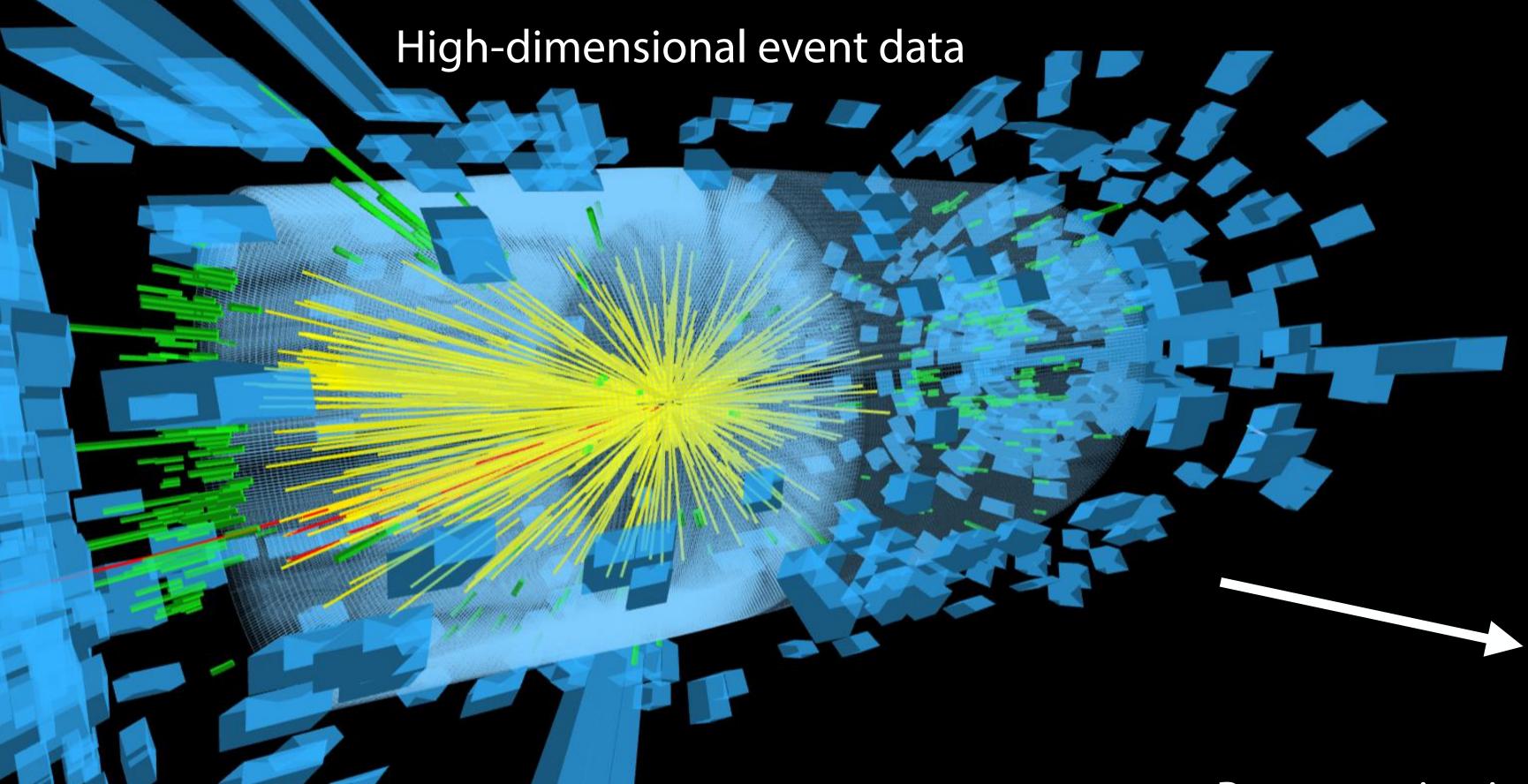
Statistical Framing



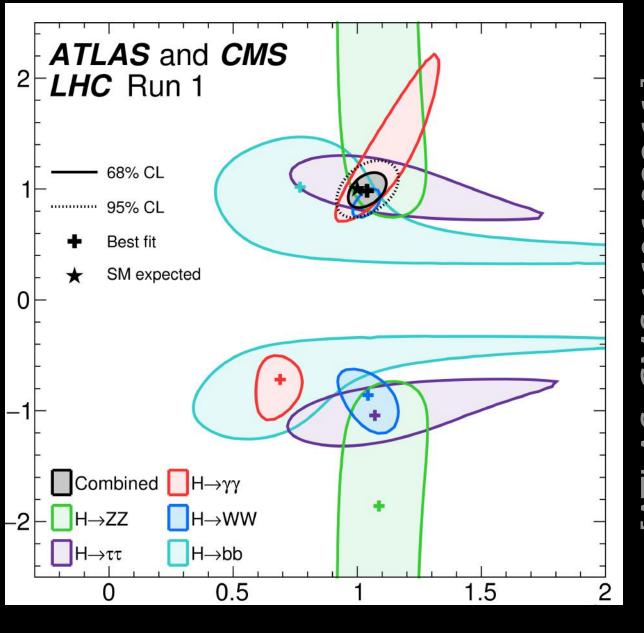




Precision constraints on new physics



[CMS]



Precision constraints on new physics

systematic expansion of

new physics around

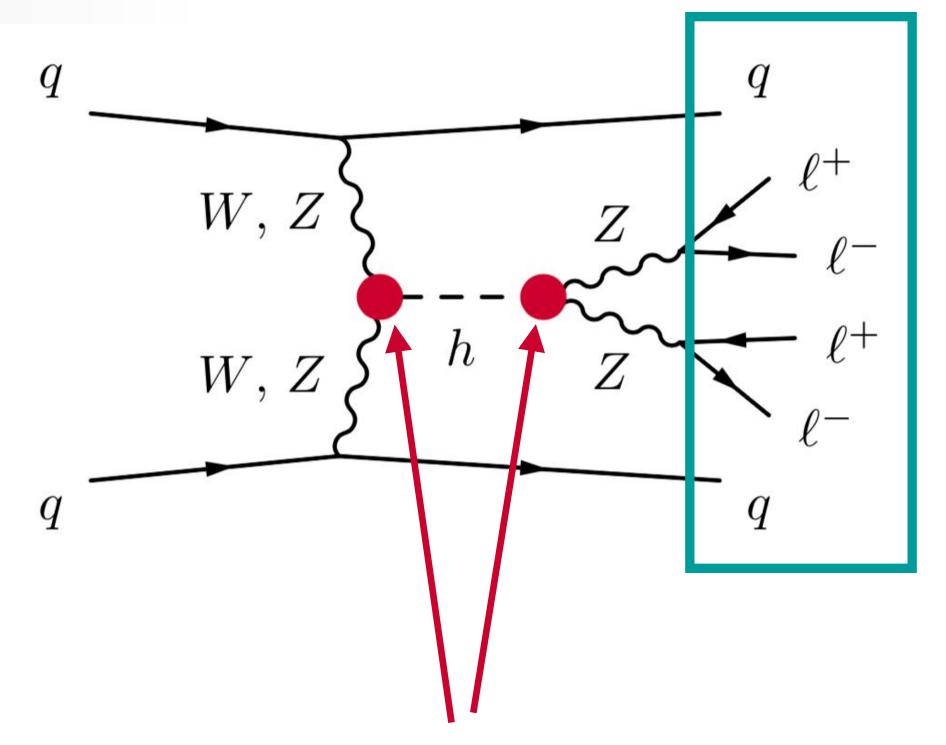
Standard Model

Parameterization e.g. in Effective Field Theory:

$$\mathcal{L}_{ ext{EFT}} = \mathcal{L}_{ ext{SM}} + \sum_{i} rac{f_{i}}{\Lambda^{2}} \mathcal{O}_{i}$$

10s to 100s "universal" parameters to measure

An Example: The Higgs Boson



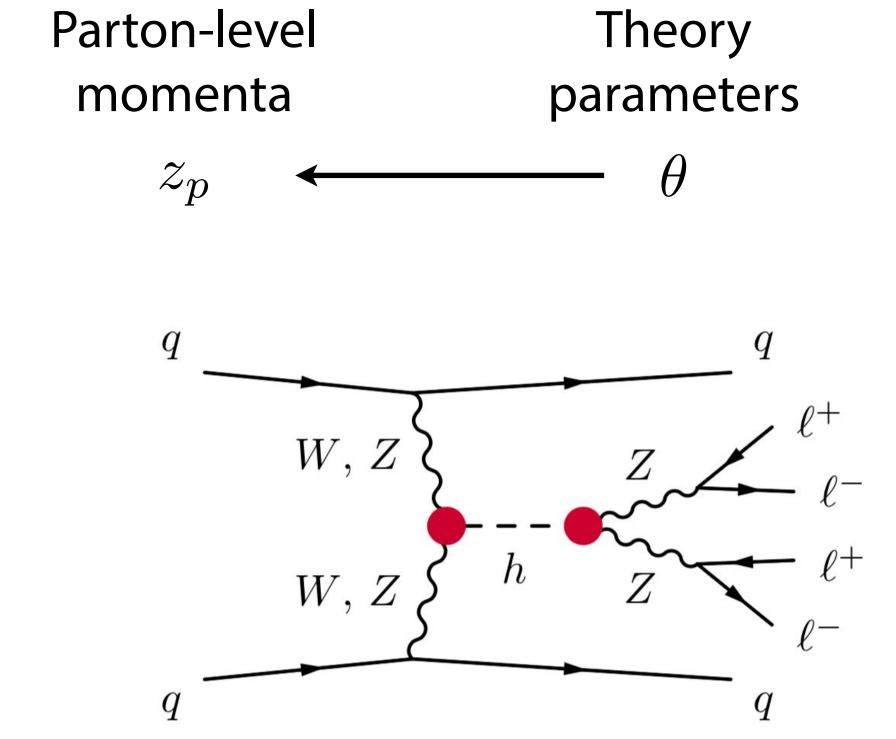
42-Dim observable x

Exciting new physics might hide here! We parameterize it with two coefficients:

$$\mathcal{L} = \mathcal{L}_{SM} + \frac{f_W}{\Lambda^2} \underbrace{\frac{ig}{2} (D^{\mu}\phi)^{\dagger} \sigma^a D^{\nu}\phi W_{\mu\nu}^a}_{\mathcal{O}_W} - \underbrace{\frac{f_{WW}}{\Lambda^2}}_{\mathcal{O}_{WW}} \underbrace{\frac{g^2}{4} (\phi^{\dagger}\phi) W_{\mu\nu}^a W^{\mu\nu}}_{\mathcal{O}_{WW}}$$

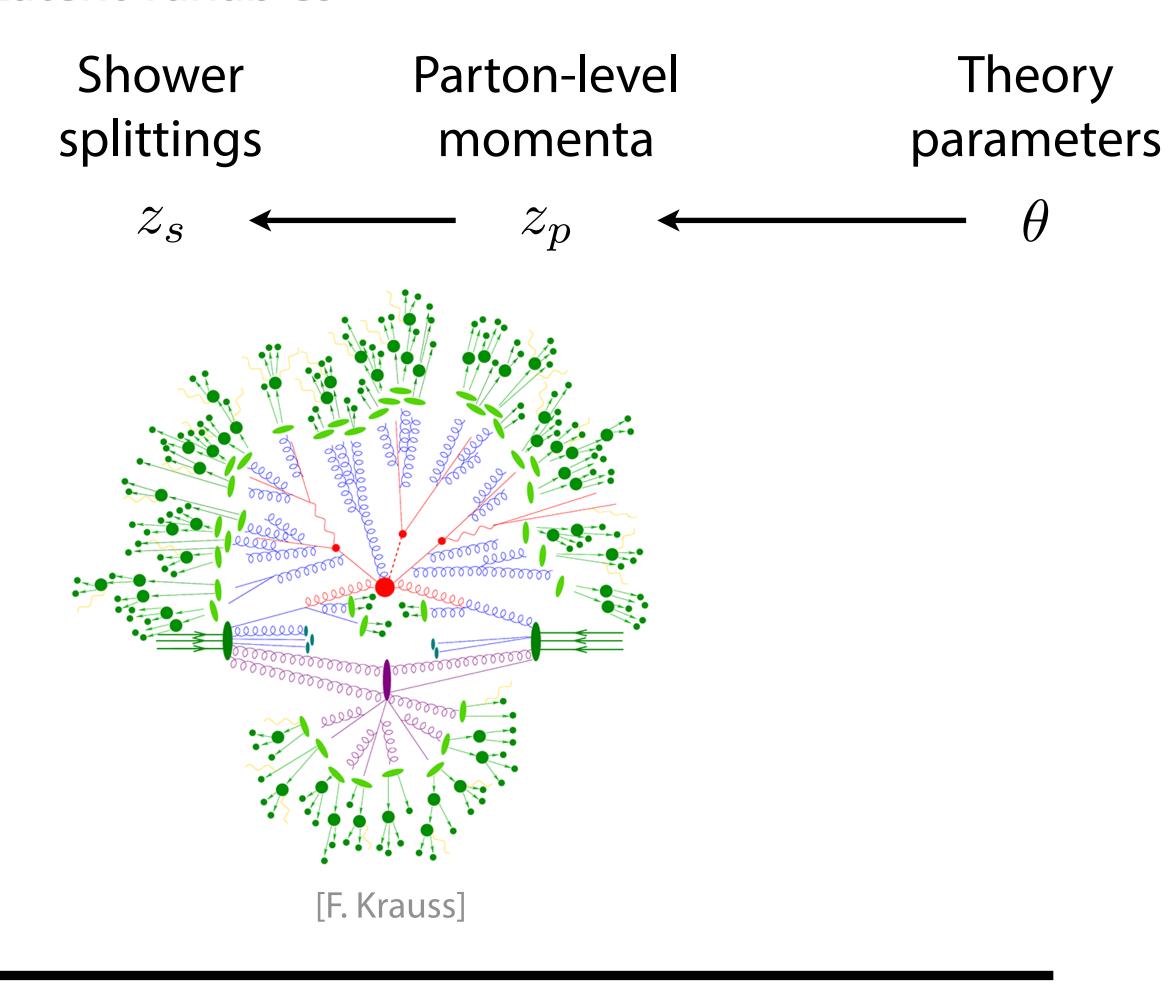
Theory parameters θ

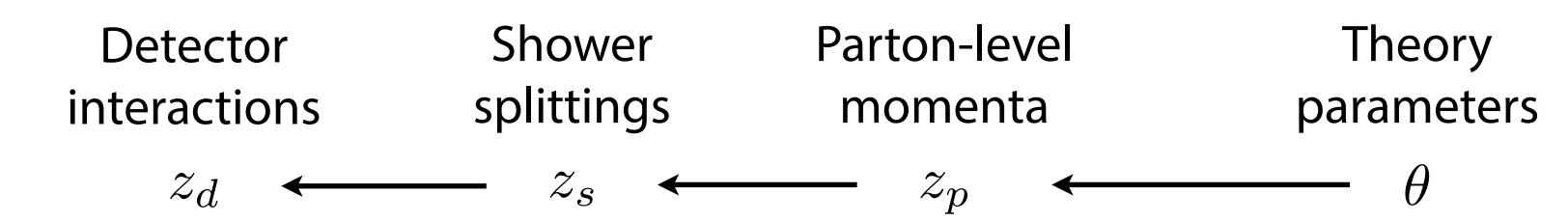
Evolution

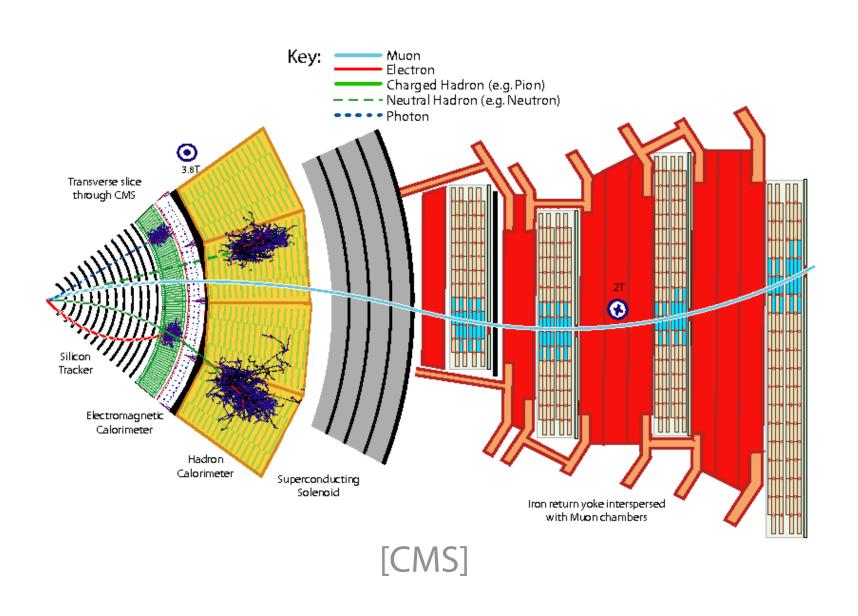


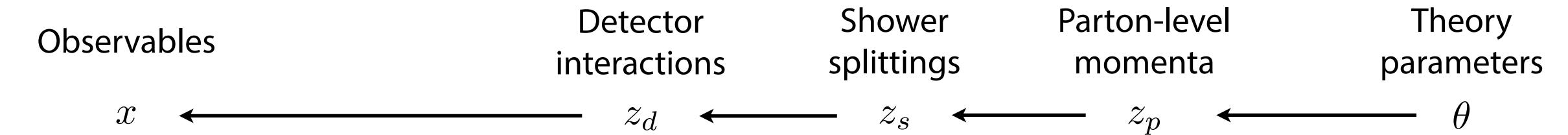


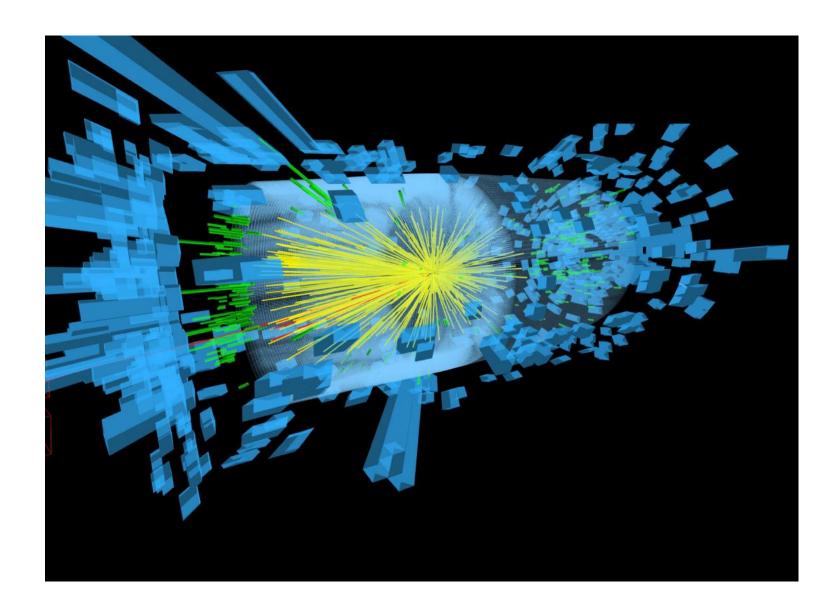


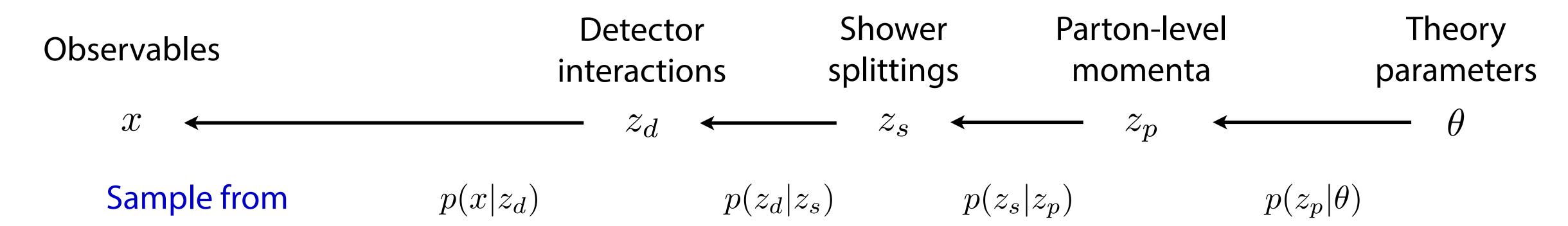






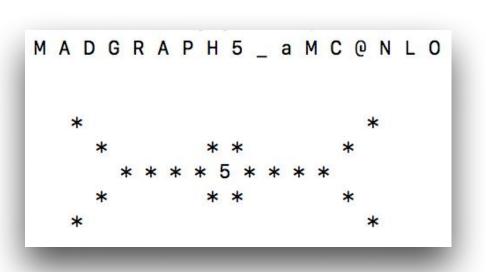


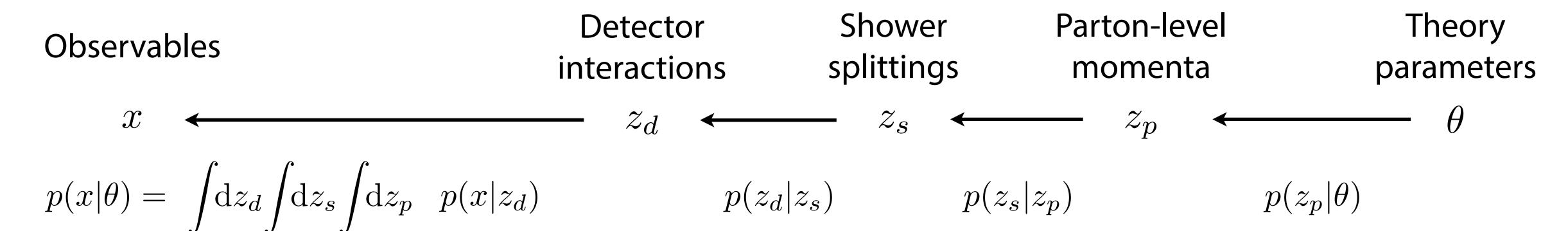




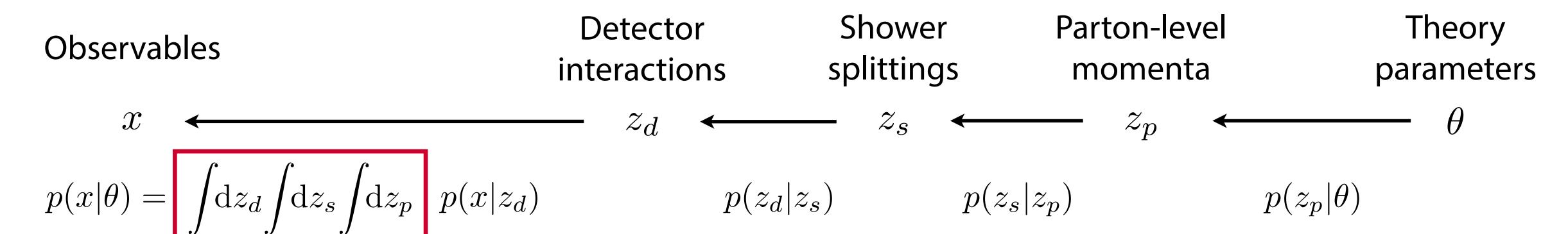








Latent variables



It's infeasible to calculate the integral over this enormous space!

Inference with intractable likelihoods

This motivates a class of inference methods for a stochastic simulator where

- evaluating the likelihood is intractable, but
- it is possible to sample synthetic data $x \sim p(x \mid \theta)$

This setting is often referred to as **likelihood-free inference**, but I prefer the term **simulation-based inference** because usually one approximates the likelihood (or likelihood ratio) and then use established inference techniques

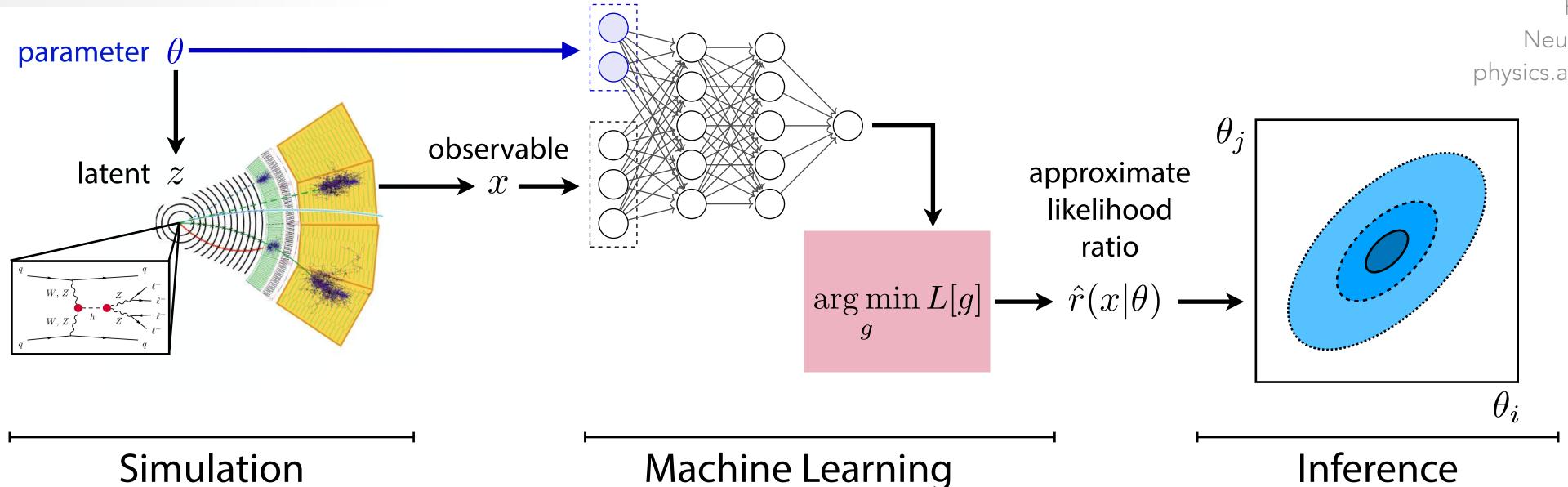
applies to both Bayesian or Frequentist inference

PNAS, arXiv:1805.12244

PRL, arXiv:1805.00013 PRD, arXiv:1805.00020

NeurlPS, arXiv:1808.00973

physics.aps.org/articles/v11/90



The surrogate for the likelihood ratio used for inference

A 2-stage process:

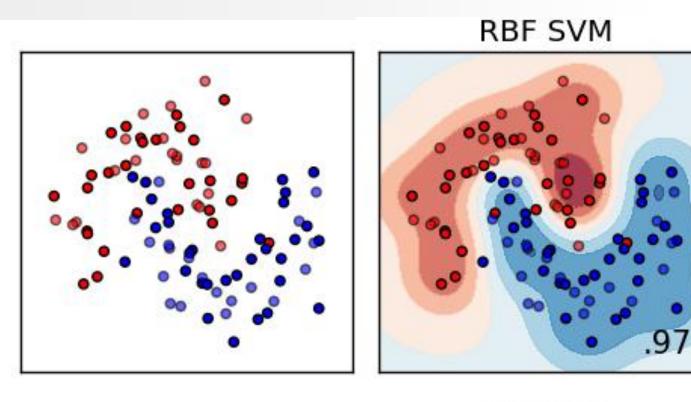
learning surrogate (amortized)

Learning the likelihood ratio

Inference on parameters of simulator (frequentist or Bayesian)

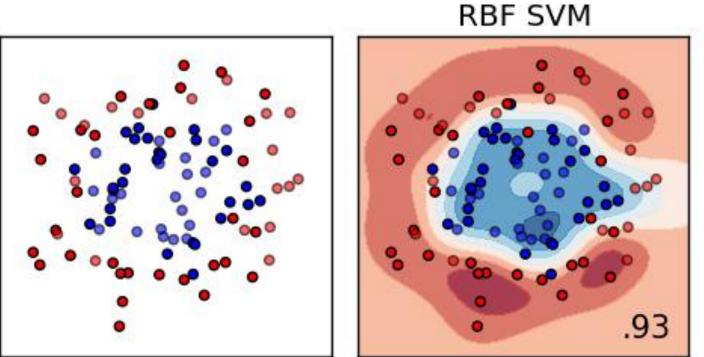
No Bayesian prior used for training, but one can use prior for inference.

Likelihood Ratio Trick



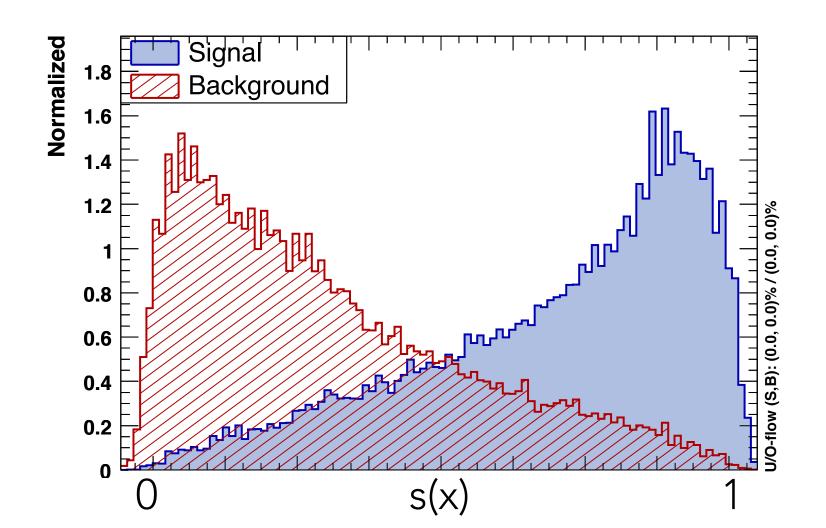
• binary classifier: find function s(x) that minimizes loss:

$$L[s] = \mathbb{E}_{p(x|H_1)}[-\log s(x)] + \mathbb{E}_{p(x|H_0)}[-\log(1 - s(x))]$$



• i.e. approximate the optimal classifier

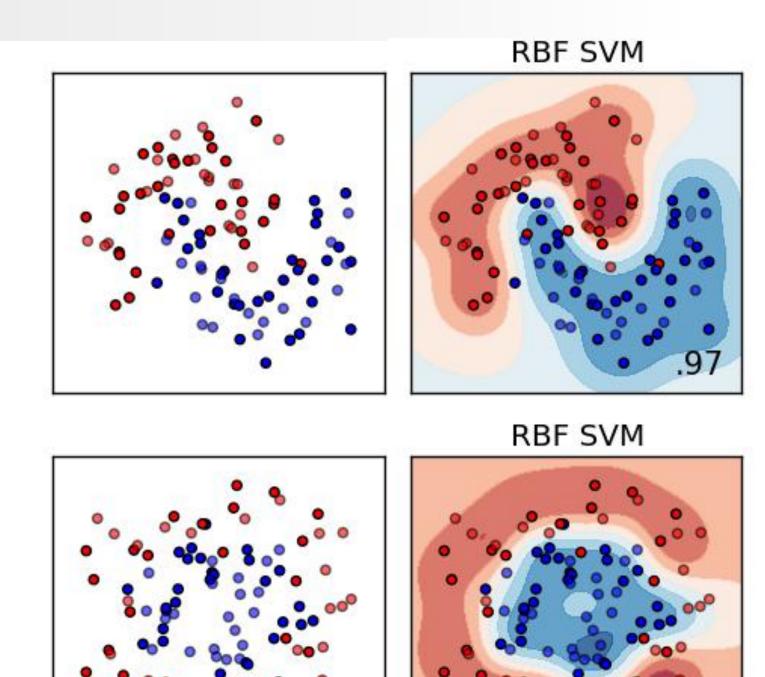
$$s(x) = \frac{p(x|H_1)}{p(x|H_0) + p(x|H_1)}$$

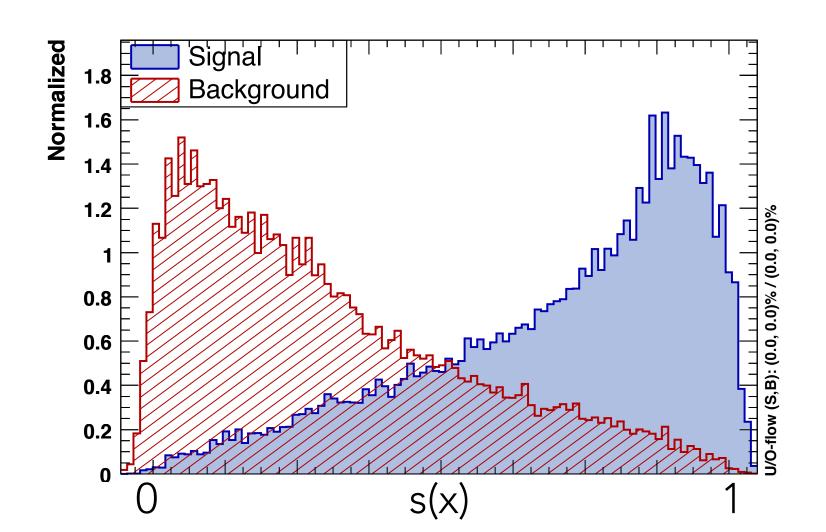


which is 1-to-1 with the likelihood ratio

$$r(x) = \frac{p(x|H_1)}{p(x|H_0)} = 1 - \frac{1}{s(x)}$$

Likelihood Ratio Trick





.93

• binary classifier: find function s(x) that minimizes loss:

$$L[s] = \mathbb{E}_{p(x|H_1)}[-\log s(x)] + \mathbb{E}_{p(x|H_0)}[-\log(1 - s(x))]$$

$$\approx \frac{1}{N} \sum_{i=1}^{N} -y_i \log s(x_i) - (1 - y_i) \log(1 - s(x_i))$$

• i.e. approximate the optimal classifier

$$s(x) = \frac{p(x|H_1)}{p(x|H_0) + p(x|H_1)}$$

which is 1-to-1 with the likelihood ratio

$$r(x) = \frac{p(x|H_1)}{p(x|H_0)} = 1 - \frac{1}{s(x)}$$

Parametrizing the Likelihood Ratio Trick

Can do the same thing for any two points $\theta_0 \& \theta_1$ in parameter space Θ .

$$r(x; \theta_0, \theta_1) = \frac{p(x \mid \theta_0)}{p(x \mid \theta_1)} = 1 - \frac{1}{s(x; \theta_0, \theta_1)}$$

Or train to classify data from $p(x \mid \theta)$ versus some fixed reference $p_{\text{ref}}(x)$

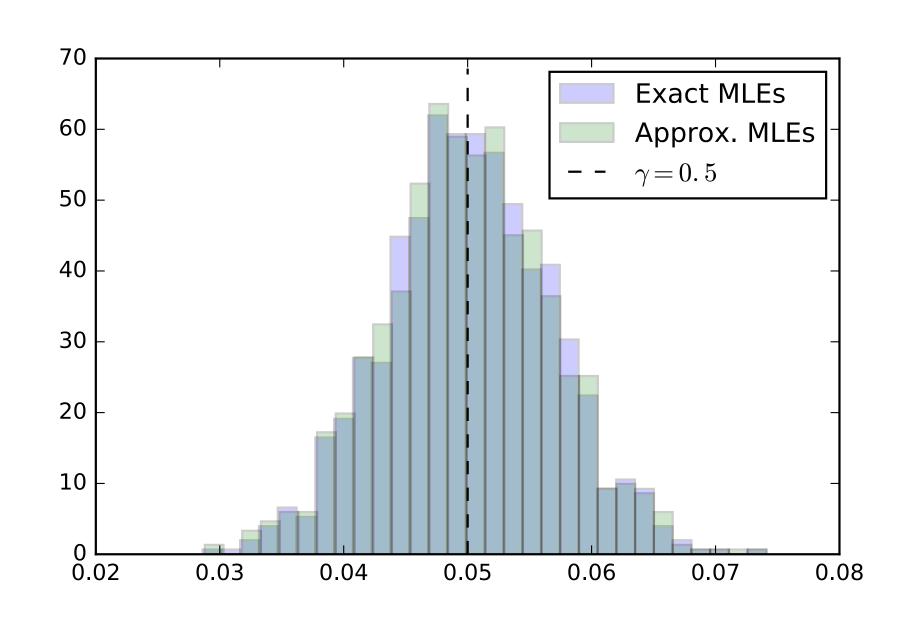
$$r(x;\theta) = \frac{p(x|\theta)}{p_{\text{ref}}(x)} = 1 - \frac{1}{s(x;\theta)}$$

I call this a parametrized classifier.

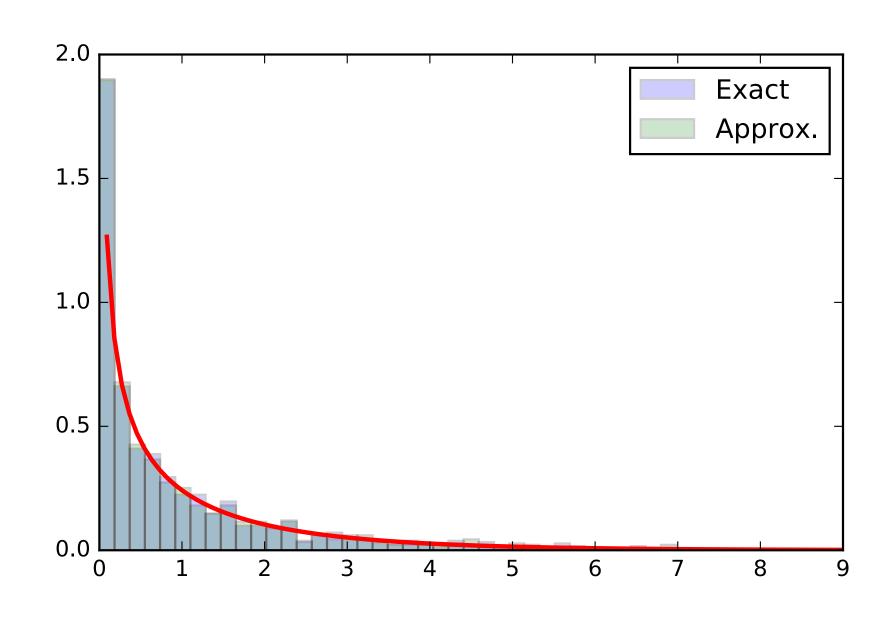
Amortized likelihood ratio

Once we've learned the likelihood ratio $r(x; \theta)$, we can apply it to any data x.

- unlike ABC, we pay biggest computational costs up front
- Great for calibrated frequentist confidence intervals with guaranteed coverage
- Here we repeat inference thousands of times & check asymptotic statistical theory



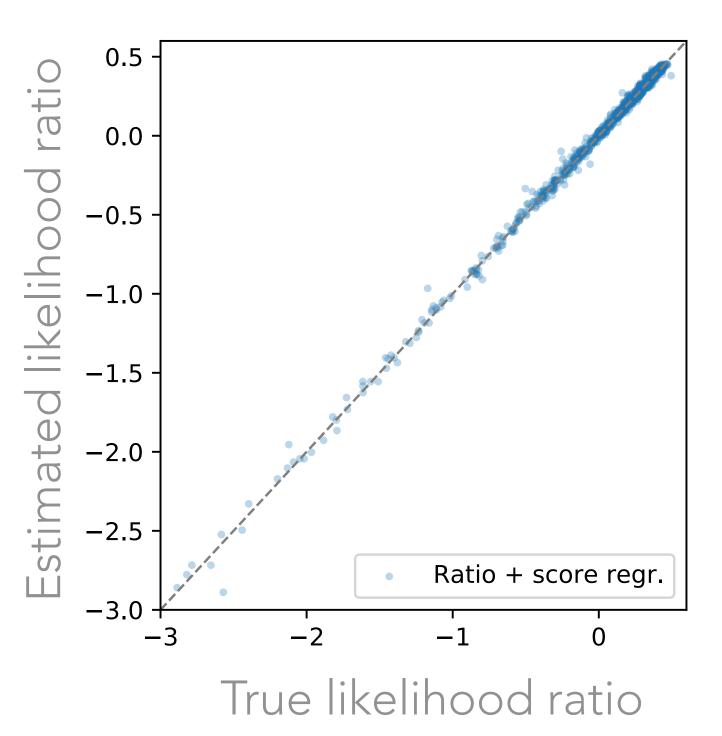
(a) Exact vs. approximated MLEs.

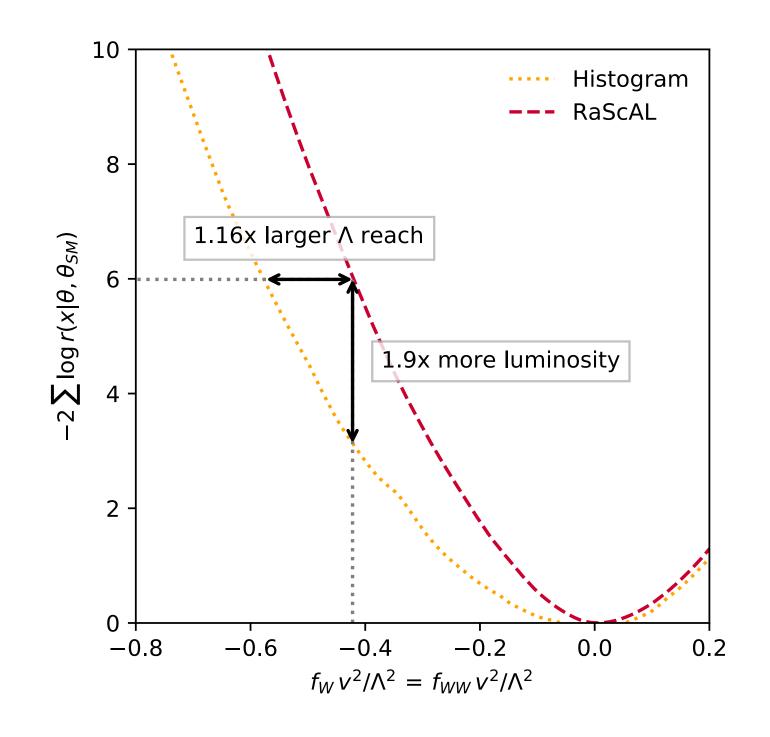


(b) $p(-2 \log \Lambda(\gamma = 0.05) | \gamma = 0.05)$

Impact on Studies of The Higgs Boson

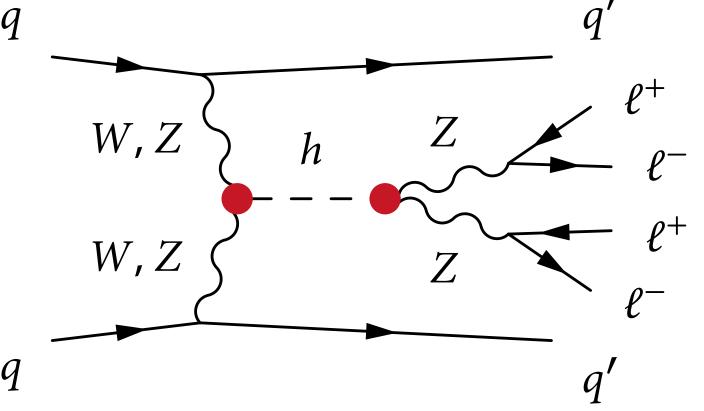
(based on a 42-Dim observation **x**)





Accurate likelihood ratio estimates without the need for summary statistics improves sensitivity significantly

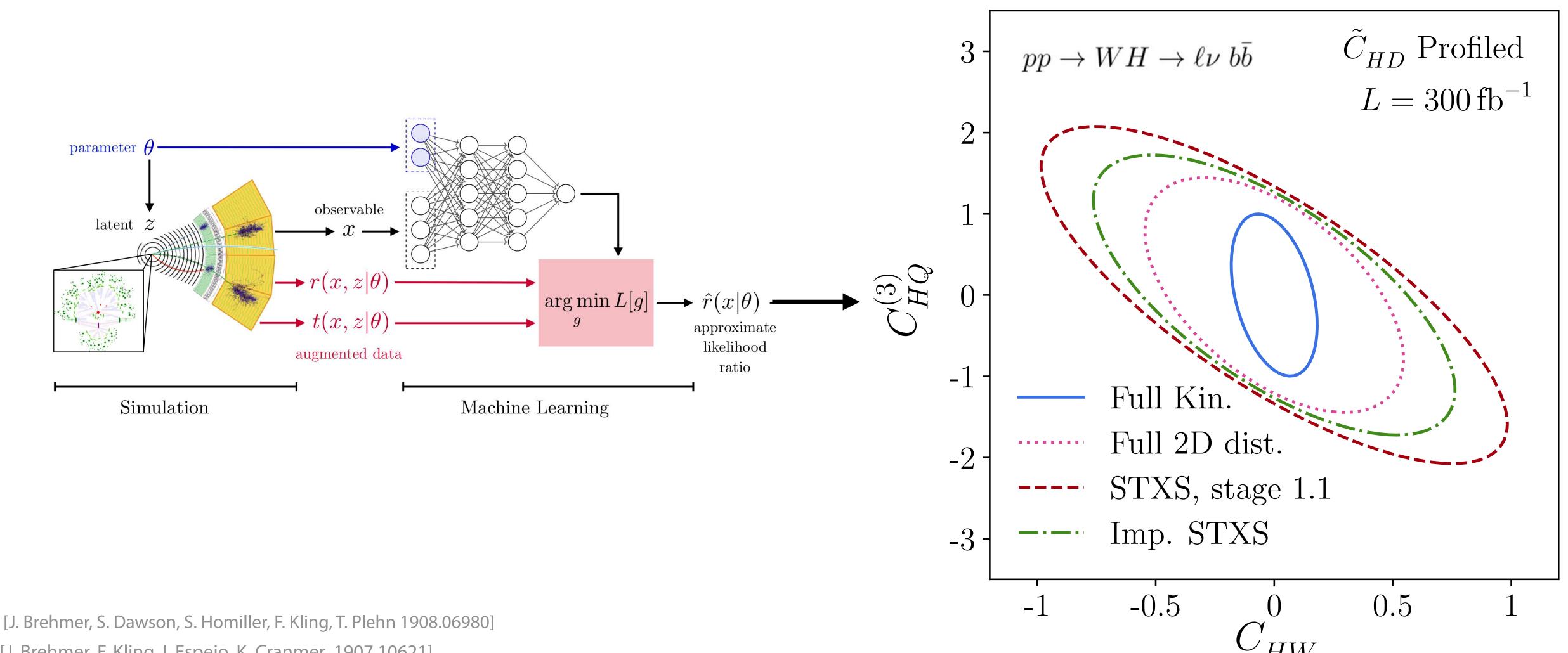
• Equivalent to 90% more LHC data!



Impact on Studies of The Higgs Boson

Massive gains in precision of a flagship measurement at the LHC!

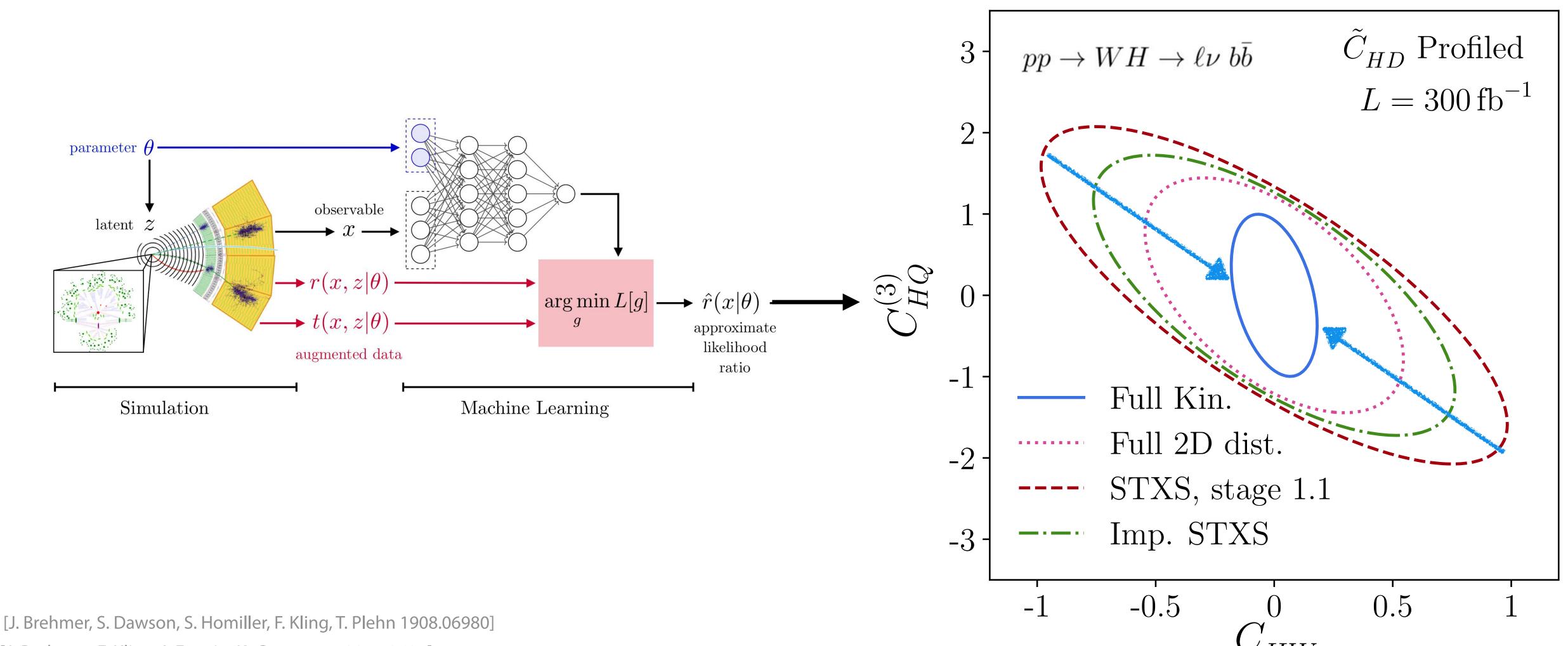
Equivalent increasing data collected by LHC by several factors



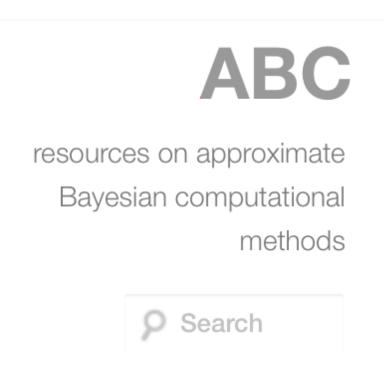
Impact on Studies of The Higgs Boson

Massive gains in precision of a flagship measurement at the LHC!

Equivalent increasing data collected by LHC by several factors



A common theme, a common language



Home

Home

This website keeps track of developments in approximate Bayesian computation (ABC) (a.k.a. likelihood-free), a class of computational statistical methods for Bayesian inference under intractable likelihoods. The site is meant to be a resource both for biologists and statisticians who want to learn more about ABC and related methods. Recent publications are under Publications 2012. A comprehensive list of publications can be found under Literature. If you are unfamiliar with ABC methods see the Introduction. Navigate using the menu to learn more.

ABC in Montreal ABC in Montreal (2014)

ABC in Montreal

Approximate Bayesian computation (ABC) or likelihood-free (LF) methods have developed mostly beyond the radar of the machine learning community, but are important tools for a large and diverse segment of the scientific community. This is particularly true for systems and population biology, computational neuroscience, computer vision, healthcare sciences, but also many others.

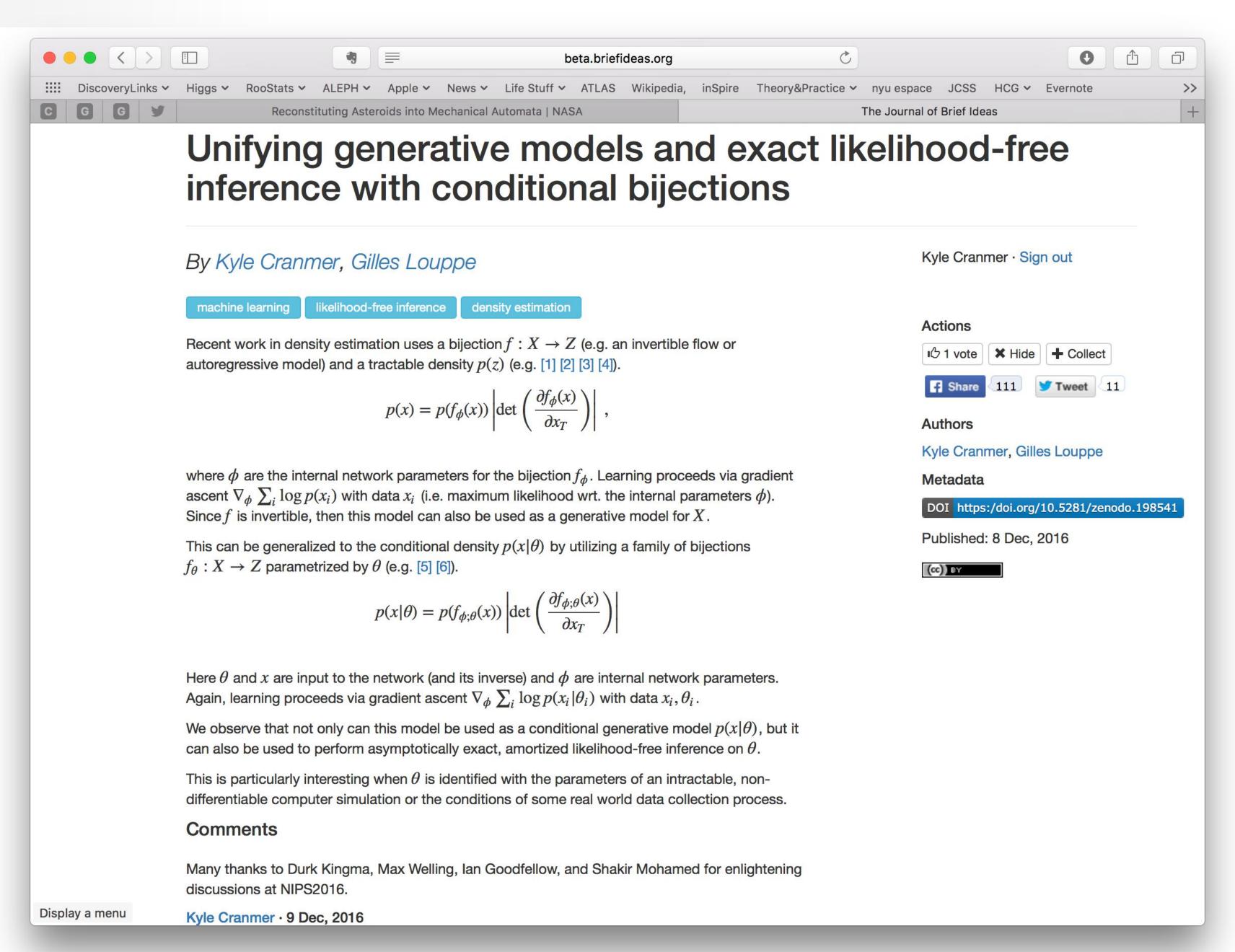
Interaction between the ABC and machine learning community has recently started and contributed to important advances. In general, however, there is still significant room for more intense interaction and collaboration. Our workshop aims at being a place for this to happen.

NIPS 2016

BARCELONA · SPAIN · DECEMBER 5 - 10, 2015 | http://nips.cc/

INVITED SPEAKERS SYMPOSIA TUTORIALS ORGANIZING COMMITTEE General Chairs: Deep Reinforcement Learning Through Policy Optimization Reproducible Research: the Case of the Human Microbiome Recurrent Neural Networks and other Daniel D Lee (University of Pennsylvania) Pieter Abbeel (OpenAl, UC Berkeley) and John Schulman (OpenAl) Susan Holmes (Stanford University) Machines that Learn Algorithms Masashi Sugiyama (The University of Tokyo) Alex Graves (Google DeepMind) Large-scale Optimization: Beyond Stochastic Gradient Dynamic Legged Robots Juergen Schmidhuber (IDSIA) Program Chairs: Descent and Convexity Marc Raibert (Boston Dynamics) Rupesh Srivastava (IDSIA) Ulrike von Luxburg (University of Tübingen) Francis Bach (INRIA, ENS) and Suvrit Sra (MIT) Sepp Hochreiter (Johannes Kepler University) Isabelle Guyon (Clopinet) Intelligent Biosphere Variational Inference: Foundations and Modern Methods Drew Purves (Google DeepMind) Deep Learning Tutorials Chair: David Blei (Columbia), Shakir Mohamed (Google Deepmind) and Navdeep Jaitly (Google) Joelle Pineau (McGill University) Roger Grosse (University of Toronto) Rajesh Ranganath (Princeton) Predictive Learning Hanna Wallach (Microsoft) Yann LeCun (Facebook and New York University) Yann LeCun (New York University & Facebook) Natural Language Processing for Computational Social Science Workshop Chairs: Machine Learning and Likelihood-Free Inference in Machine Learning and the Law Cristian Danescu-Niculescu-Mizil (Cornell) and Lillian Lee (Cornell) Ralf Herbrich (Amazon) Adrian Weller (Cambridge, Alan Turing Inst.) Particle Physics Kyle Cranmer (New York University) Conrad McDonnell (Gray's Inn Tax Chambers) Generative Adversarial Networks Demonstration Chair: Jatinder Singh (University of Cambridge) Ian Goodfellow (OpenAI) Raia Hadsell (Google DeepMind) Learning About the Brain: Neuroimaging and Beyond Thomas Grant (University of Cambridge) Theory and Algorithms for Forecasting Non-stationary Time Series Irina Rish (IBM T.J. Watson Research Center) Publications Chair & Electronic Proceedings Chair: Vitaly Kuznetsov (Google) and Mehryar Mohri (Courant Institute, Roman Garnett (Washington University) Engineering Principles From Stable And Developing Google Research) Program Managers: Saket Navlakha (The Salk Institute for Biological Studie Deep Learning for Building Al Systems Krikamol Muandet (Mahidol University and MPI) Andrew Ng (Baidu, Stanford University) Rohit Babbar, Behzad Tabibian (MPI for Intelligent Systems) ML Foundations and Methods for Precision Medicine and Healthcare Suchi Saria (Johns Hopkins) and Peter Schulam (Johns Hopkin Crowdsourcing: Beyond Label Generation Jenn Wortman Vaughan (Microsoft Research)

Invertible NNs / Normalizing Flows



ICML 2017 Workshop on Implicit Models

Workshop Aims

Probabilistic models are an important tool in machine learning. They form the basis for models that generate realistic data, uncover hidden structure, and make predictions. Traditionally, probabilistic models in machine learning have focused on prescribed models. Prescribed models specify a joint density over observed and hidden variables that can be easily evaluated. The requirement of a tractable density simplifies their learning but limits their flexibility --- several real world phenomena are better described by simulators that do not admit a tractable density. Probabilistic models defined only via the simulations they produce are called implicit models.

Arguably starting with generative adversarial networks, research on implicit models in machine learning has exploded in recent years. This workshop's aim is to foster a discussion around the recent developments and future directions of implicit models.

Implicit models have many applications. They are used in ecology where models simulate animal populations over time; they are used in phylogeny, where simulations produce hypothetical ancestry trees; they are used in physics to generate particle simulations for high energy processes. Recently, implicit models have been used to improve the state-of-the-art in image and content generation. Part of the workshop's focus is to discuss the commonalities among applications of implicit models.

Of particular interest at this workshop is to unite fields that work on implicit models. For example:

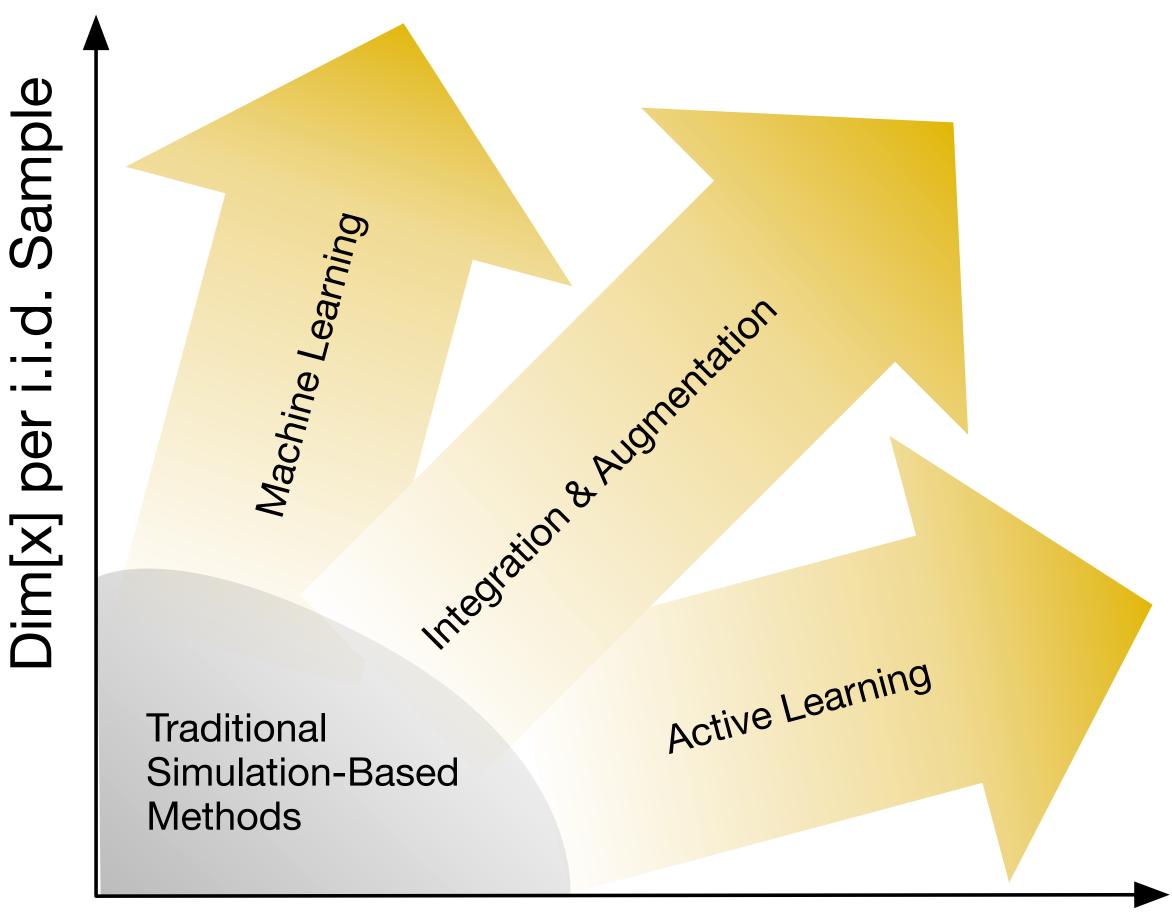
- **Generative adversarial networks** (a NIPS 2016 workshop) are implicit models with an adversarial training scheme.
- Recent advances in variational inference (a NIPS 2015 and 2016 workshop) have leveraged implicit models for more accurate approximations.
- Approximate Bayesian computation (a NIPS 2015 workshop) focuses on posterior inference for models with implicit likelihoods.
- Learning implicit models is deeply connected to two sample testing, density ratio and density difference estimation.

We hope to bring together these different views on implicit models, identifying their core challenges and combining their innovations.

The frontier of simulation-based inference

Kyle Cranmer^{a,b,1}, Johann Brehmer^{a,b}, and Gilles Louppe^c

^aCenter for Cosmology and Particle Physics, New York University, USA; ^bCenter for Data Science, New York University, USA; ^cMontefiore Institute, University of Liège, Belgium April 3, 2020



Gilles Louppe



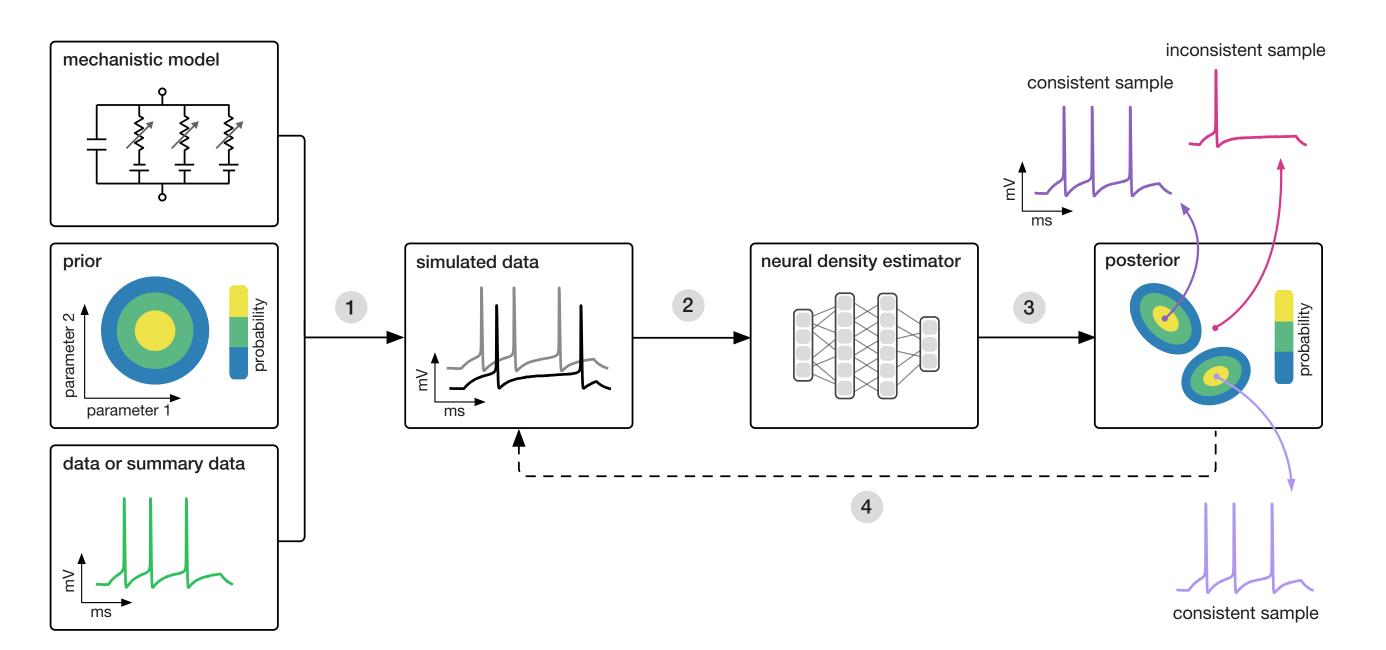
Johann Brehmer

In computational neuroscience



sbi: A toolkit for simulation-based inference

Alvaro Tejero-Cantero^{e, 1}, Jan Boelts^{e, 1}, Michael Deistler^{e, 1}, Jan-Matthis Lueckmann^{e, 1}, Conor Durkan^{e, 2}, Pedro J. Gonçalves^{1, 3}, David S. Greenberg^{1, 4}, and Jakob H. Macke^{1, 5, 6}



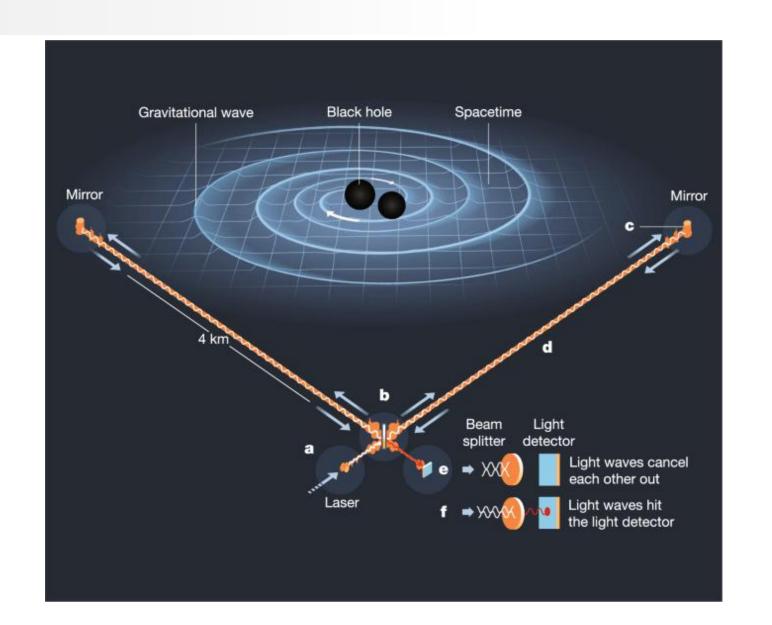
Training deep neural density estimators to identify mechanistic models of neural dynamics

Pedro J Gonçalves^{1,2†*}, Jan-Matthis Lueckmann^{1,2†*}, Michael Deistler^{1,3†*}, Marcel Nonnenmacher^{1,2,4}, Kaan Öcal^{2,5}, Giacomo Bassetto^{1,2}, Chaitanya Chintaluri^{6,7}, William F Podlaski⁶, Sara A Haddad⁸, Tim P Vogels^{6,7}, David S Greenberg^{1,4}, Jakob H Macke^{1,2,3,9*}

¹Computational Neuroengineering, Department of Electrical and Computer Engineering, Technical University of Munich, Munich, Germany; ²Max Planck Research Group Neural Systems Analysis, Center of Advanced European Studies and Research (caesar), Bonn, Germany; ³Machine Learning in Science, Excellence Cluster Machine Learning, Tübingen University, Tübingen, Germany; ⁴Model-Driven Machine Learning, Institute of Coastal Research, Helmholtz Centre Geesthacht, Geesthacht, Germany; ⁵Mathematical Institute, University of Bonn, Bonn, Germany; ⁶Centre for Neural Circuits and Behaviour, University of Oxford, Oxford, United Kingdom; ⁷Institute of Science and Technology Austria, Klosterneuburg, Austria; ⁸Max Planck Institute for Brain Research, Frankfurt, Germany; ⁹Max Planck Institute for Intelligent Systems, Tübingen, Germany

Abstract Mechanistic modeling in neuroscience aims to explain observed phenomena in terms of underlying causes. However, determining which model parameters agree with complex and stochastic neural data presents a significant challenge. We address this challenge with a machine learning tool which uses deep neural density estimators—trained using model simulations—to carry out Bayesian inference and retrieve the full space of parameters compatible with raw data or selected data features. Our method is scalable in parameters and data features and can rapidly analyze new data after initial training. We demonstrate the power and flexibility of our approach on receptive fields, ion channels, and Hodgkin–Huxley models. We also characterize the space of circuit configurations giving rise to rhythmic activity in the crustacean stomatogastric ganglion, and use these results to derive hypotheses for underlying compensation mechanisms. Our approach will help close the gap between data-driven and theory-driven models of neural dynamics.

In gravitational wave astronomy



Real-time gravitational-wave science with neural posterior estimation

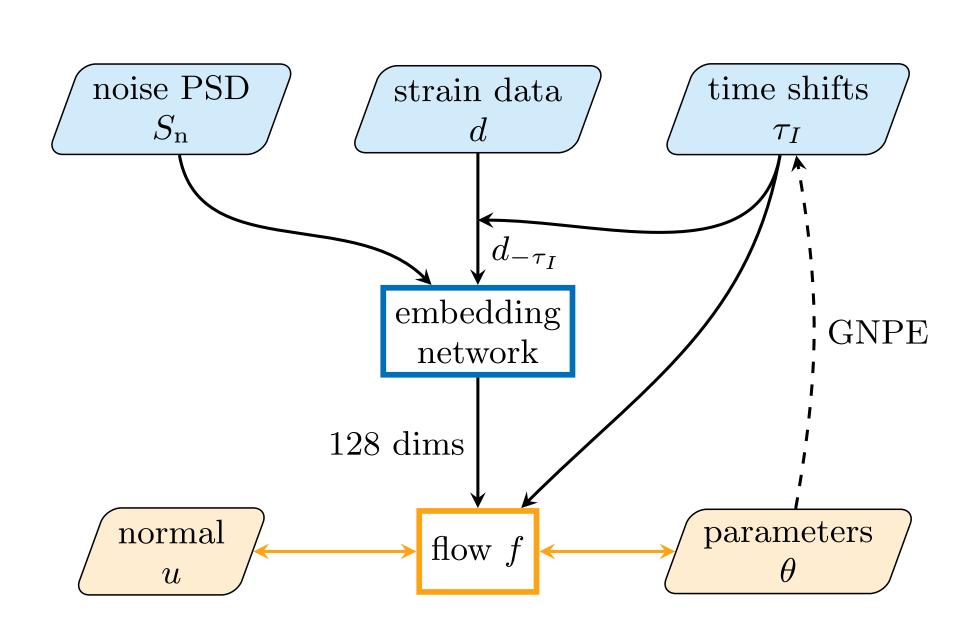
Maximilian Dax,^{1,*} Stephen R. Green,^{2,†} Jonathan Gair,^{2,‡} Jakob H. Macke,^{1,3} Alessandra Buonanno,^{2,4} and Bernhard Schölkopf¹

¹Max Planck Institute for Intelligent Systems, Max-Planck-Ring 4, 72076 Tübingen, Germany

²Max Planck Institute for Gravitational Physics (Albert Einstein Institute), Am Mühlenberg 1, 14476 Potsdam, Germany

³Machine Learning in Science, University of Tübingen, 72076 Tübingen, Germany

⁴Department of Physics, University of Maryland, College Park, MD 20742, USA



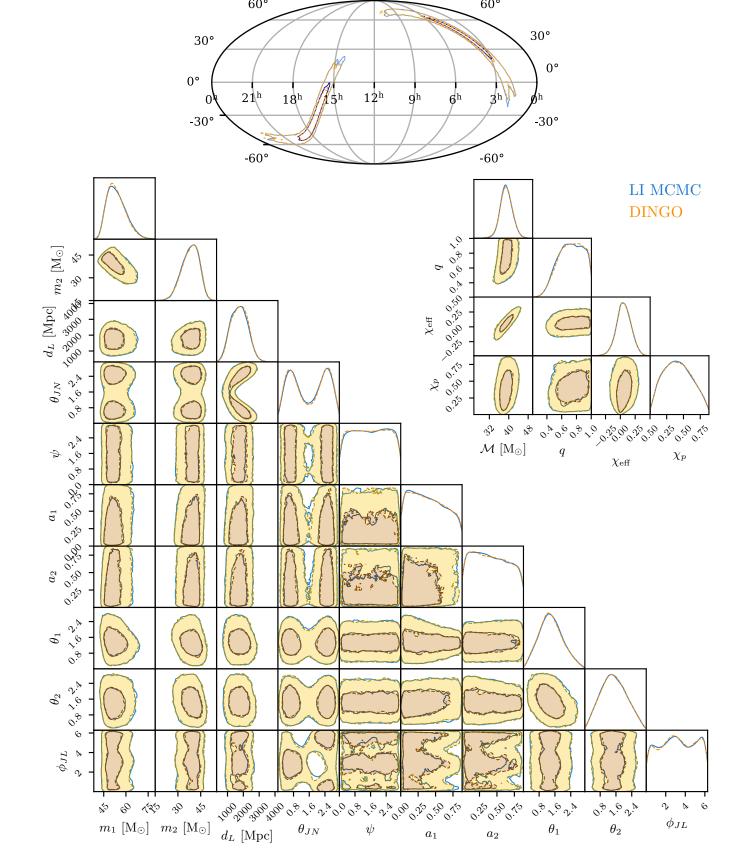


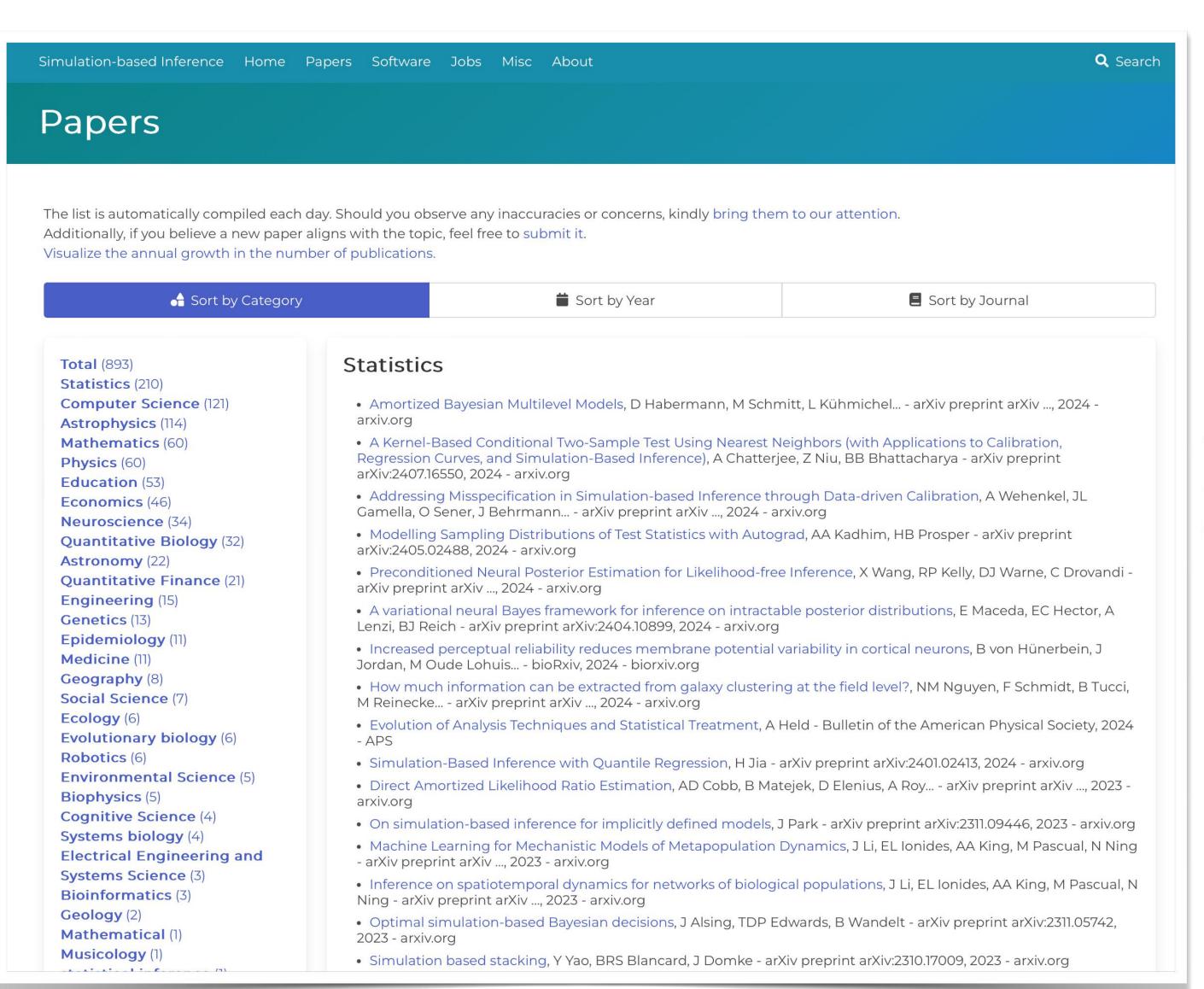
Figure 15. GW170823.



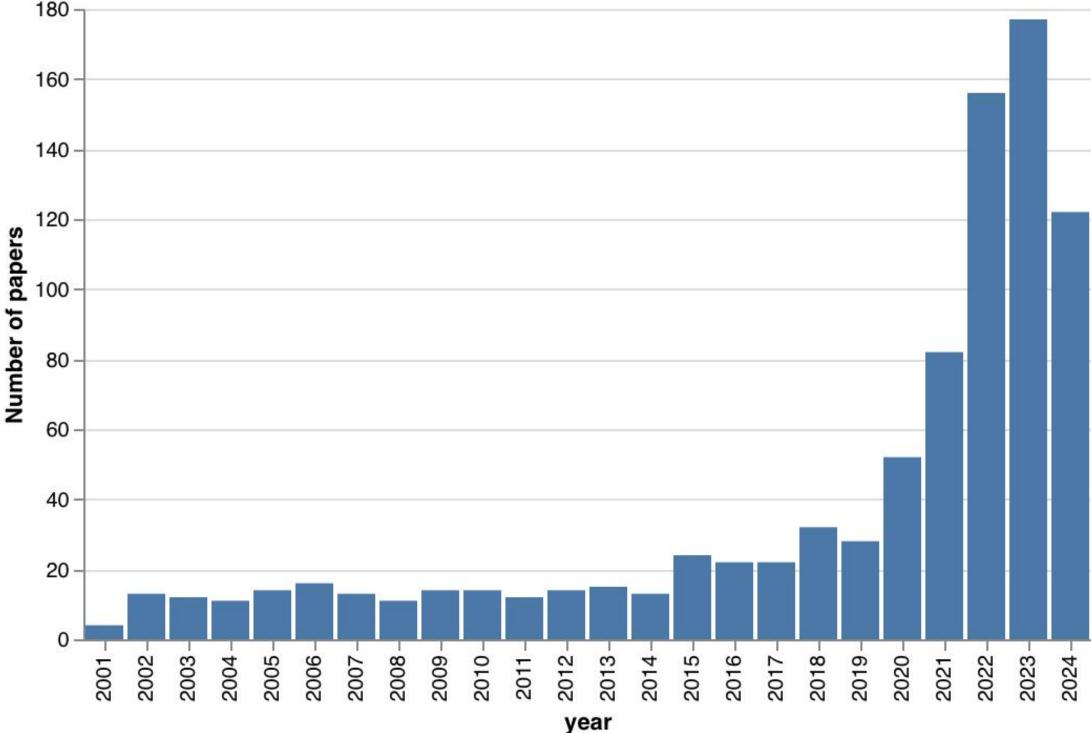
Earlier today @ STAMPS



simulation-based-inference.org



Number of Simulation-based Inference Papers by Year



github.com/smsharma/awesome-neural-sbi



Awesome Neural SBI



A community-sourced list of papers and resources on neural simulation-based inference, covering both methodological developments and domain applications. Given the nature of the field, the list is bound to be highly incomplete -- contributions are welcome!

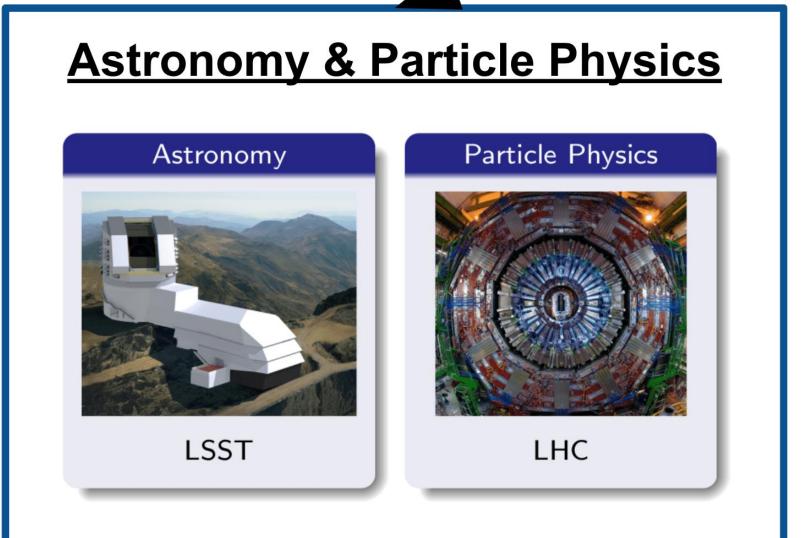
Contents

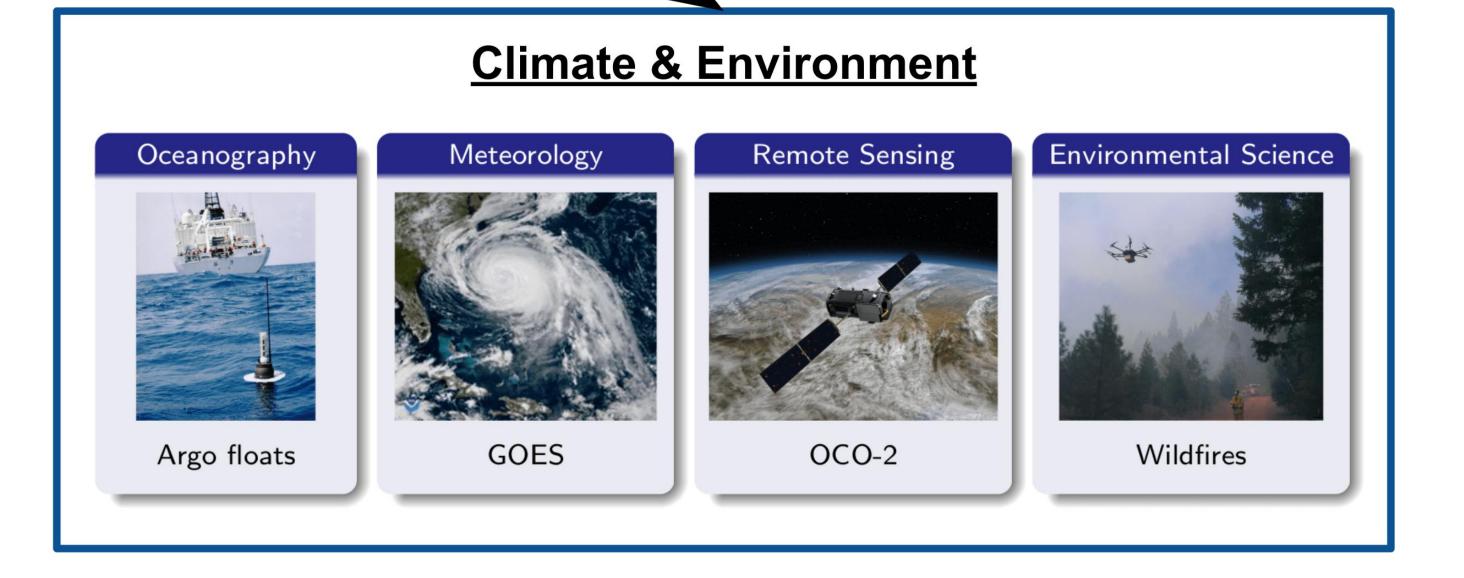
- Software and Resources
 - Code Packages and Benchmarks
 - Review Papers
 - Discovery and Links
- Papers: Methods
- Papers: Application
 - Cosmology, Astrophysics, and Astronomy
 - Particle Physics
 - Neuroscience
 - Health and Medicine
 - Other Domains
 - Application to Real Data

Focus areas for STAMPS

Foundational Methodology

Statistics
Data Science
Machine Learning / Al

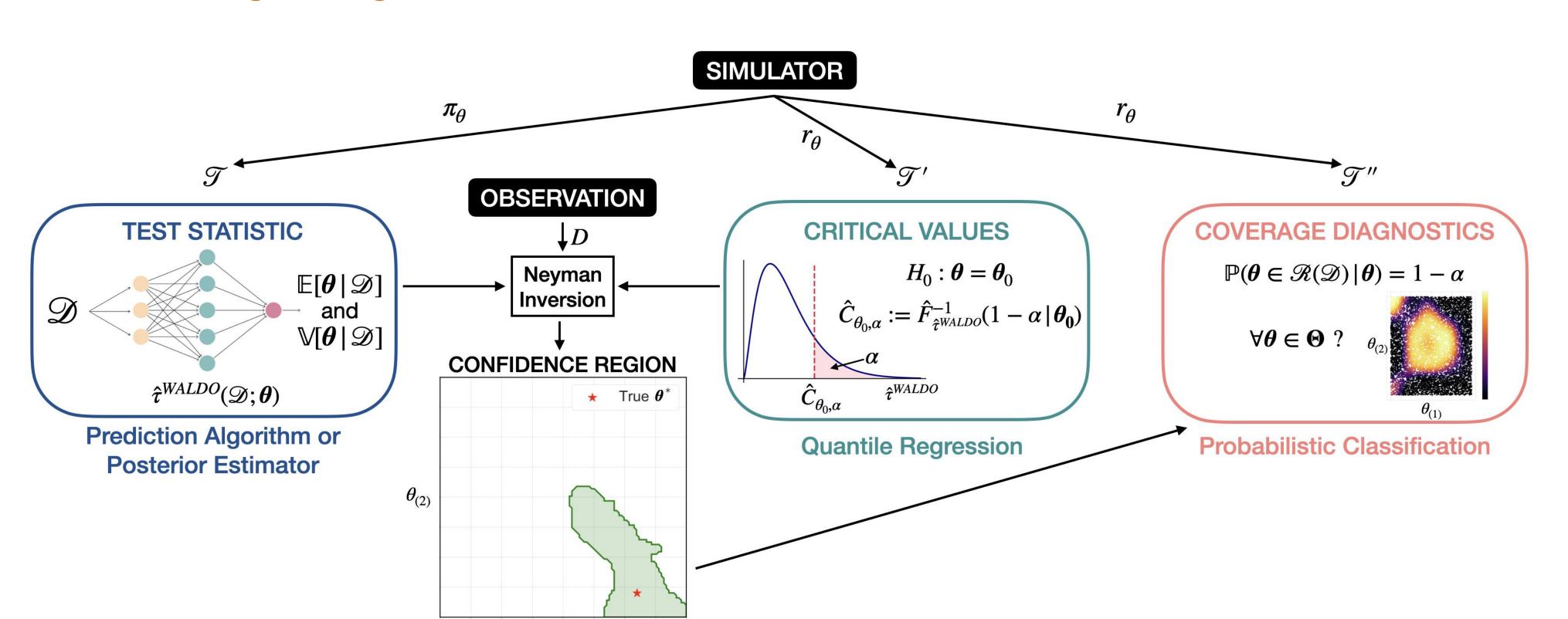




Advancing Frequentist SBI

Taking a modular approach, revisiting core issues in frequentist inference in SBI context

- Choice of test statistic for composit tests with nuisance parameters
- Efficient estimate of critical values
- Coverage diagnostics, etc.





Ann B. Lee



Luca Masserano

SBI for spatial statistics

SBI for spatial statistics

"Is there a more general notion of SBI?" \rightarrow Yes!

In recent years, there has been an explosion of interest in SBI / neural inference / amortized inference for purely statistical models especially in spatial statistics:

- Neural prediction for spatial models (Gerber and Nychka, 2021; Lenzi et al., 2023; Sainsbury-Dale et al., 2024)
- Neural likelihood for spatial models (Walchessen et al., 2024)
- Neural prediction with censored observations (Richards et al., 2023)
- Neural prediction with irregularly spaced observations (Sainsbury-Dale et al., 2023)

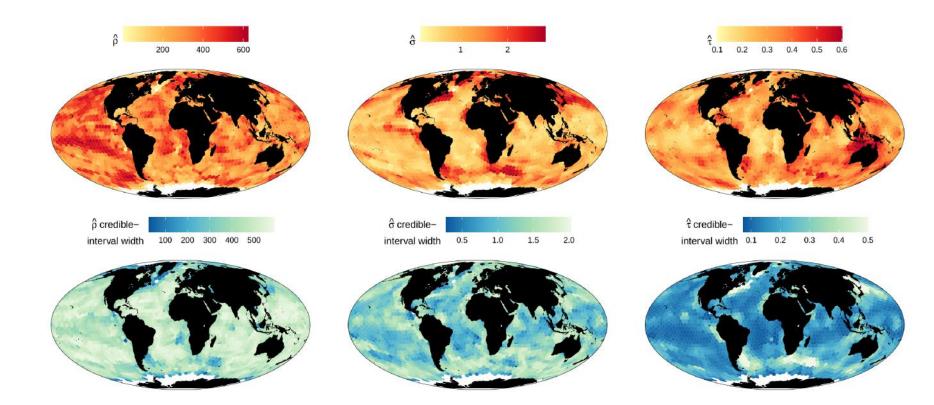


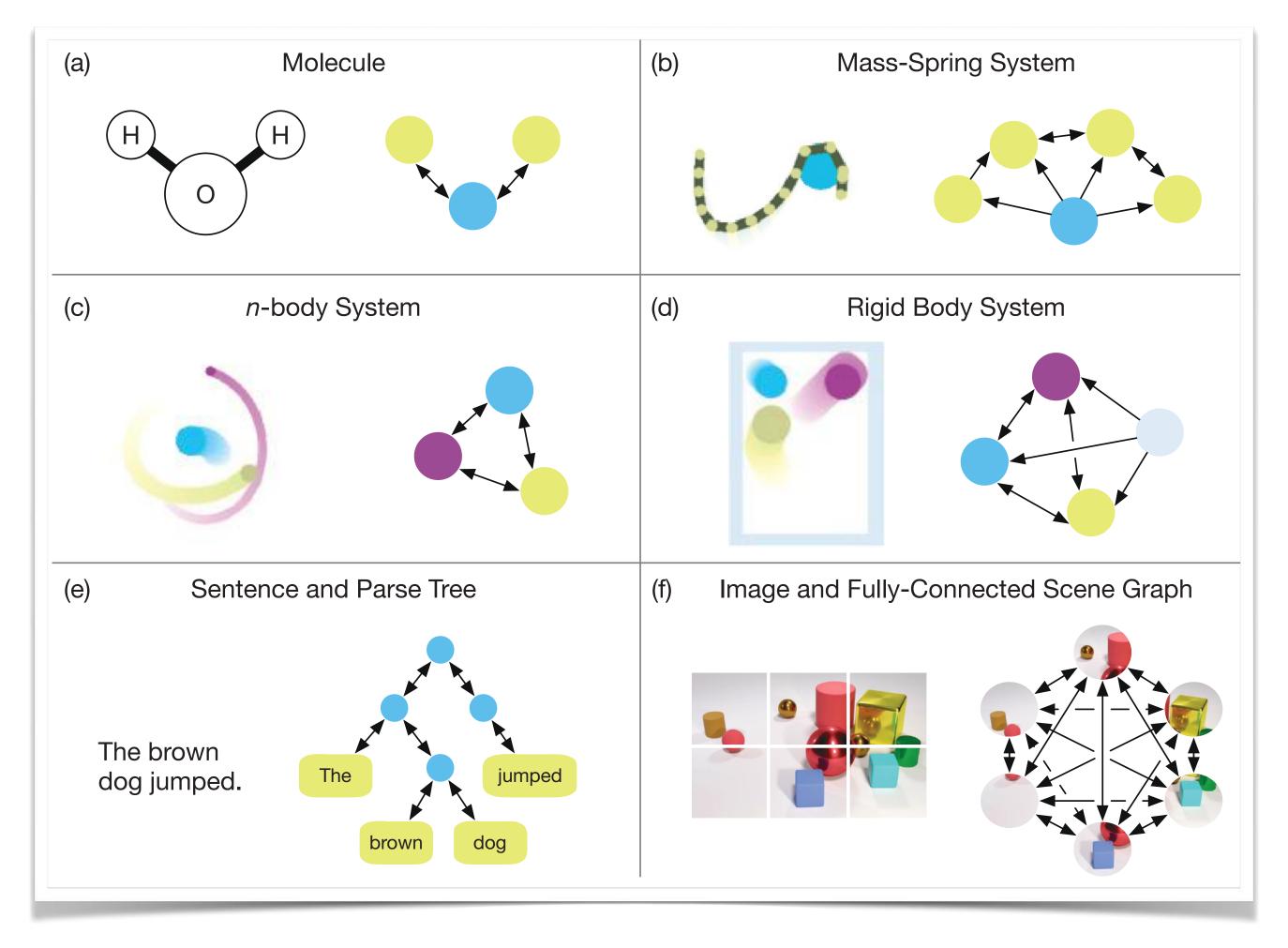
Figure: Neural prediction for satellite sea surface temperature using locally stationary Gaussian processes (Sainsbury-Dale et al., 2023)

Mikael Kuusela

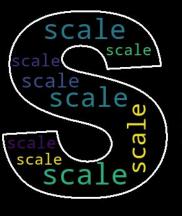
Mikael Kuusela (CMU) September 12, 2024 16 / 16

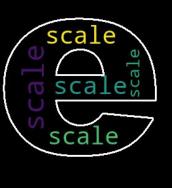
Incorporating Domain Knowledge into ML Models

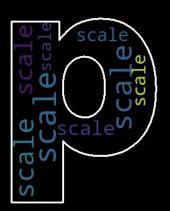
Insight of data generating process informs inductive bias on architecture

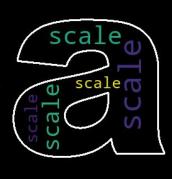


Inductive Bias Compositionality Relationships Symmetry Causality

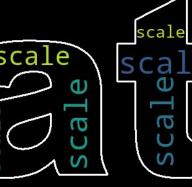






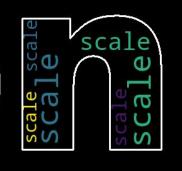


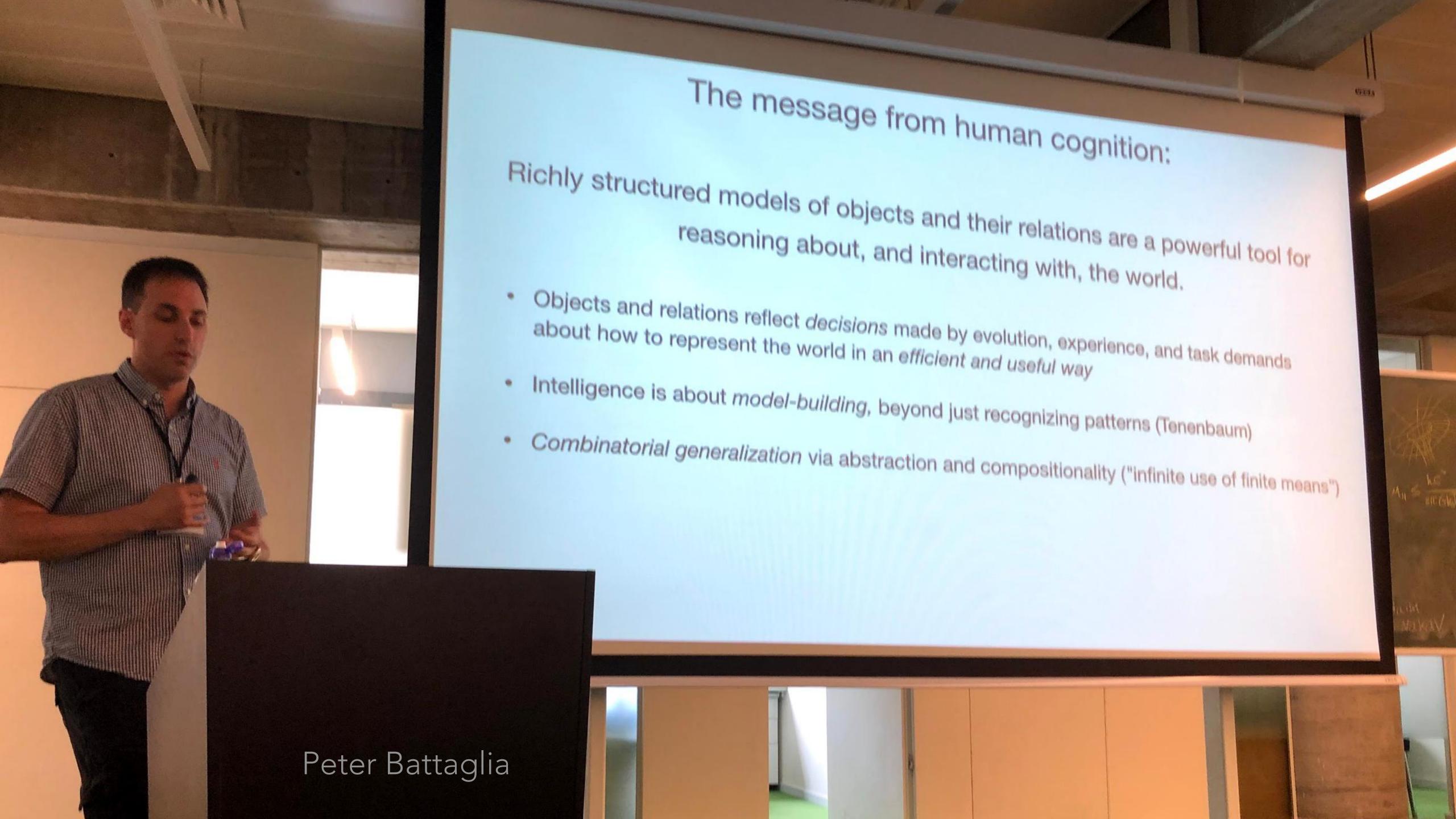


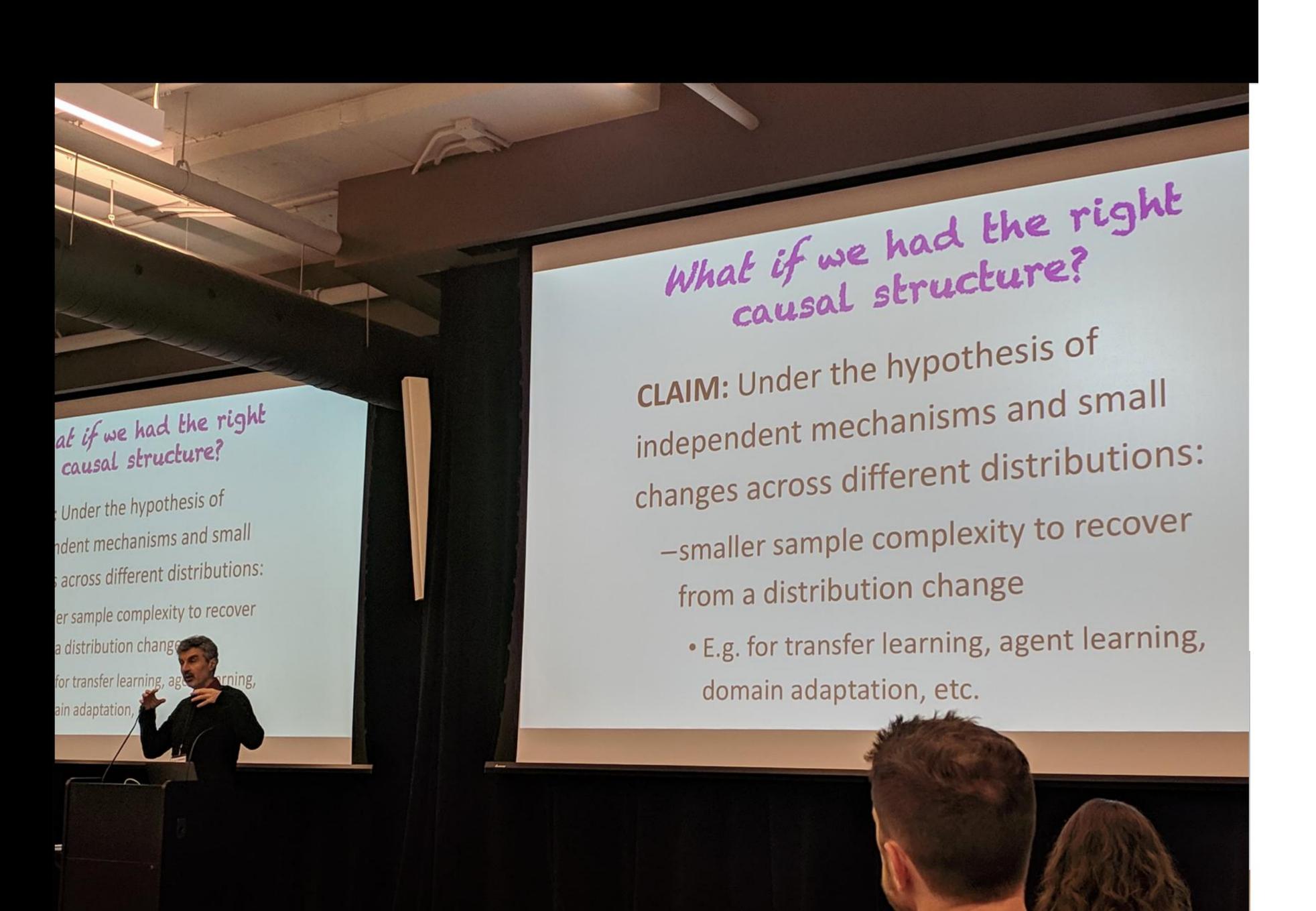














Max Welling Isn't this what Bernhard Schoelkopf has been saying for a while?

Like · Reply · 6w



Yann LeCun ...and Leon Bottou?

Like · Reply · 6w



Leon Bottou Yoshua's paper says: if you observe a distribution change that comes from a causal effect, then you'll adapt faster if your generative model matches the causal model.

Another way of seeing it is: the right causal graph suggests a particular factorization of the joint distribution (a directed bayesian network). A causal intervention means that you only change one of these factors (or a few factors) while leaving the other ones unchanged. Therefore if your generative model is the right causal model, meaning that it factorizes the joint in the same way, it will be easy to adapt it to the change because only a few parameters need changing (those associated with the factors that actually changed).



Max Welling Dan Roy I am, and I think most of us, are keenly aware that Josh has been the big proponent of this view. And I think most people agree with him on this view. Integrating this view with deep learning for more narrowly defined tasks seems to me an interesting intellectual pursuit though. I think that's what's happening here but I was not at the talk

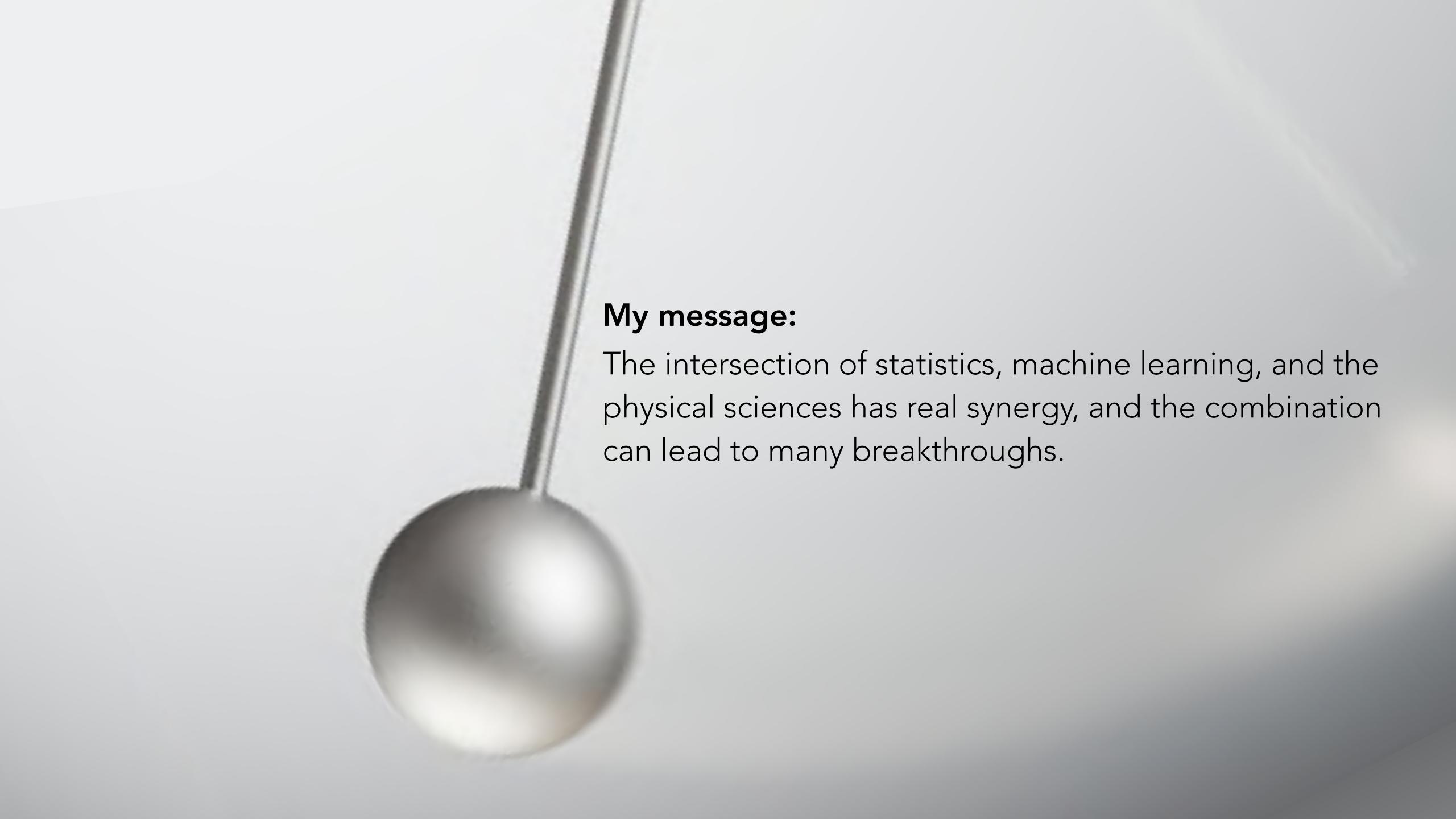
Conclusion

Human Intelligence

- Traditional scientific approaches
- largely guided by expert knowledge and theoretical insights
- hand-crafted

Artificial Intelligence

- data-driven & Big Data
- eschew expert knowledge
- end-to-end learning



Thank you.

Questions?

References

Opinionated review

K. Cranmer, J. Brehmer, G. Louppe: "The frontier of simulation-based inference" [1911.01429]

Review focusing on particle physics use-case

K. Cranmer and Johann Brehmer "Simulation-based inference methods for particle physics" [Artificial Intelligence for Particle Physics [book] and arXiv:2010.06439]

Do It Yourself (for LHC physics)

J. Brehmer, F. Kling, I. Espejo, K. Cranmer: "MadMiner: Machine learning—based inference for particle physics" [CSBS, 1907.10621, https://github.com/diana-hep/madminer]

LHC HXSWG YR4 STXS

J. Brehmer, S. Dawson, S. Homiller, F. Kling, T. Plehn: "Benchmarking simplified template cross sections in WH production" [JHEP, 1908.06980]

LHC EFT WG Report

LHC EFT Working Group "Benchmarking simplified template cross sections in WH production" [arXiv: 2211.08353]

Original works

J. Brehmer, K. Cranmer, G. Louppe, J. Pavez: "A guide to constraining Effective Field Theories with machine learning" [PRD, 1805.00020]

J. Brehmer, G. Louppe, J. Pavez, K. Cranmer: "Mining gold from implicit models to improve likelihood-free inference"

[PNAS, 1805.12244]

Follow-up with incremental improvements

M. Stoye, J. Brehmer, K. Cranmer, G. Louppe, J. Pavez:

"Likelihood-free inference with an improved cross-entropy estimator"

[NeurlPS workshop, 1808.00973]