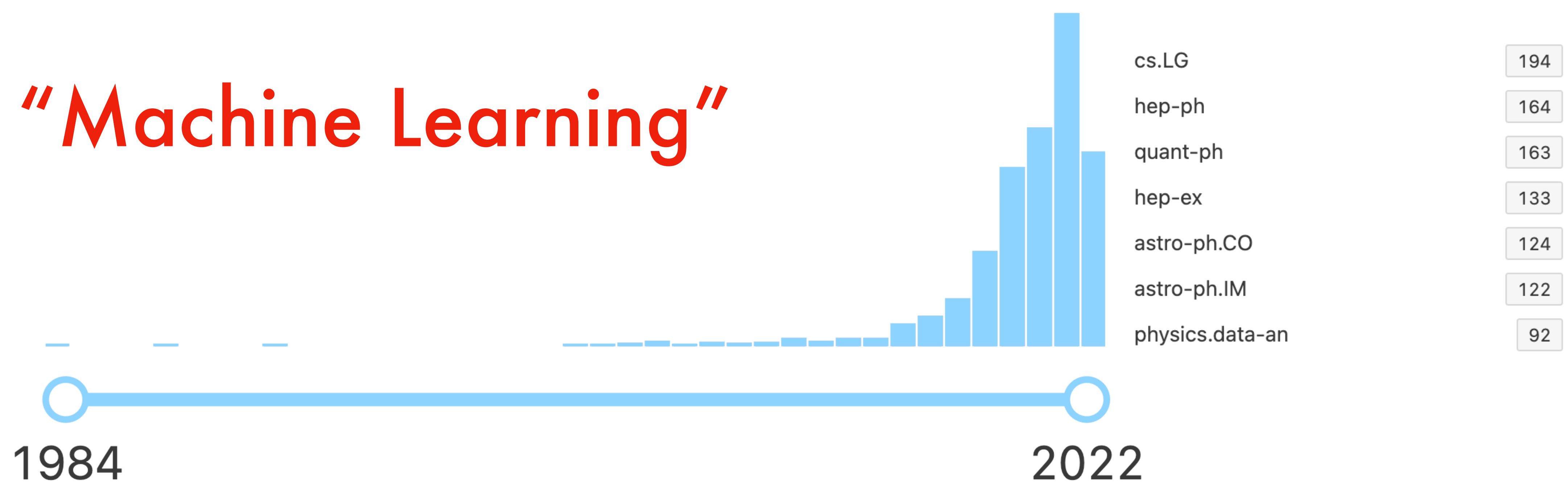
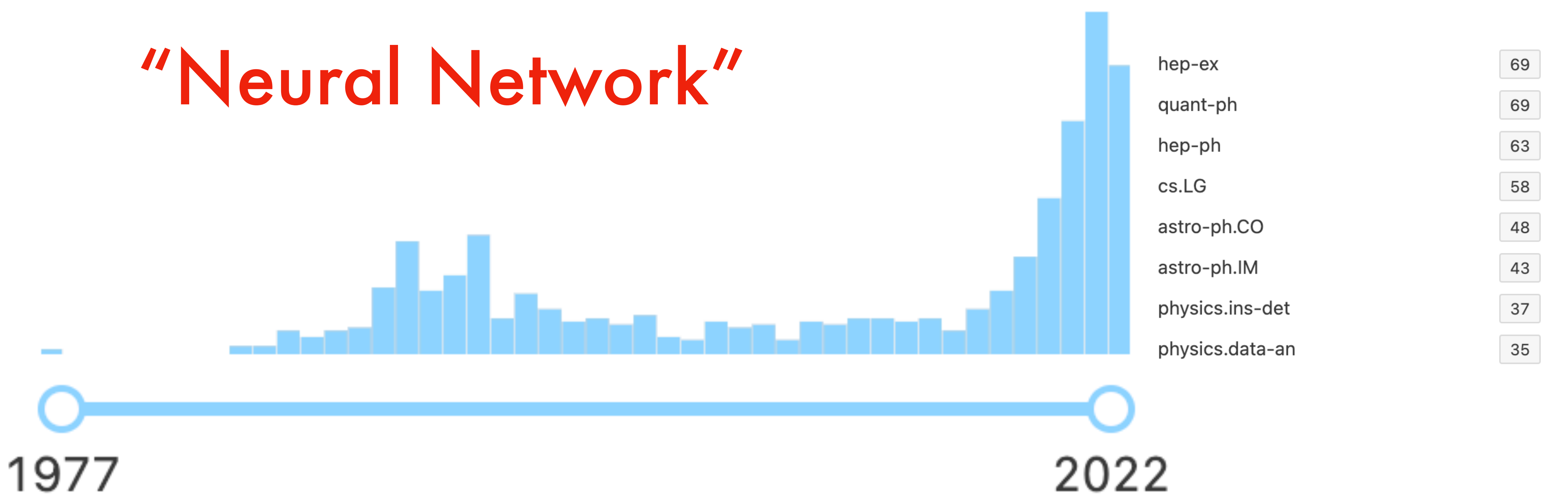
Learning Physics From Machines

Daniel Whiteson, UC Irvine

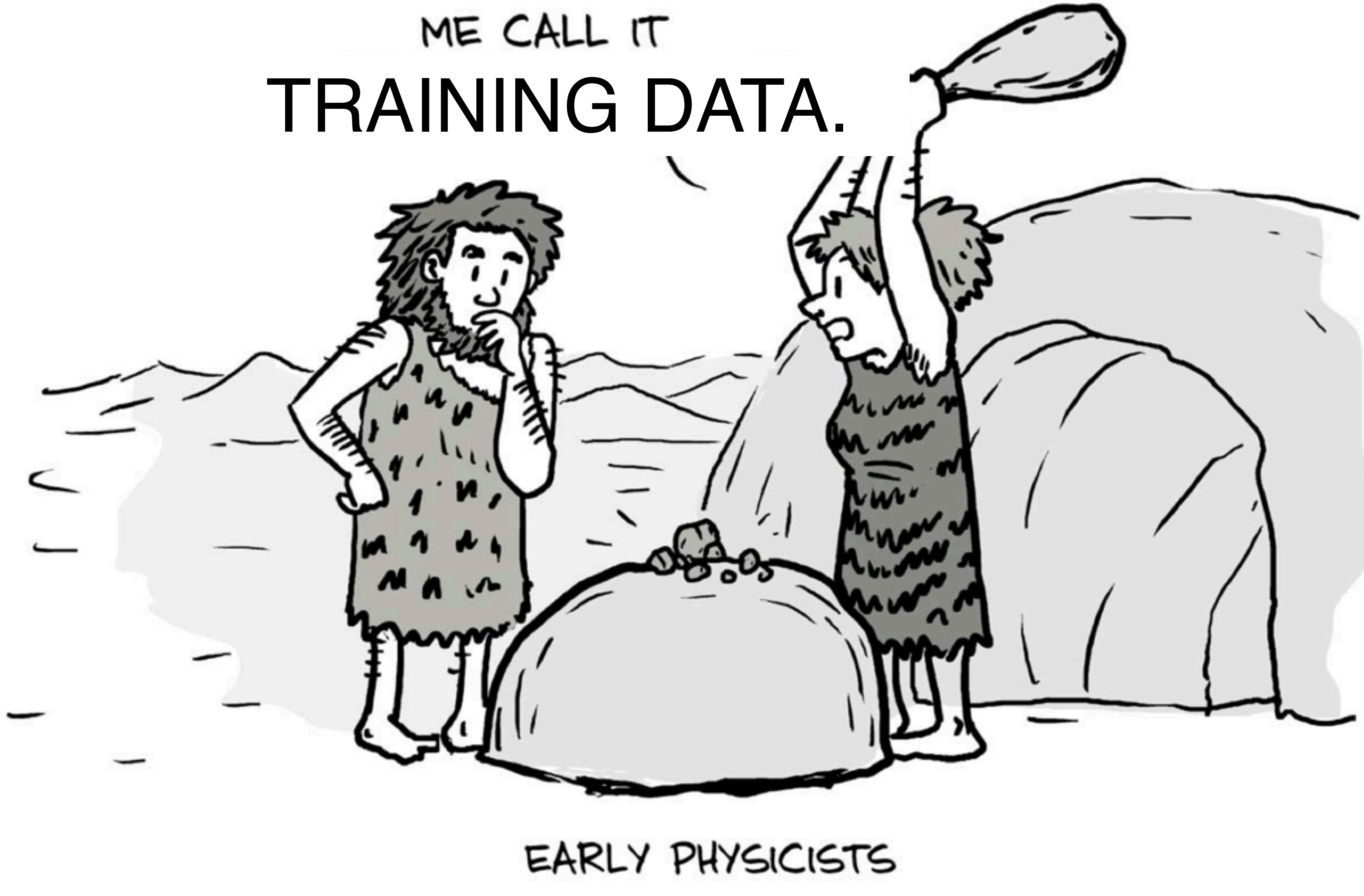
Sep 2023

It's everywhere!





First days of ML in physics



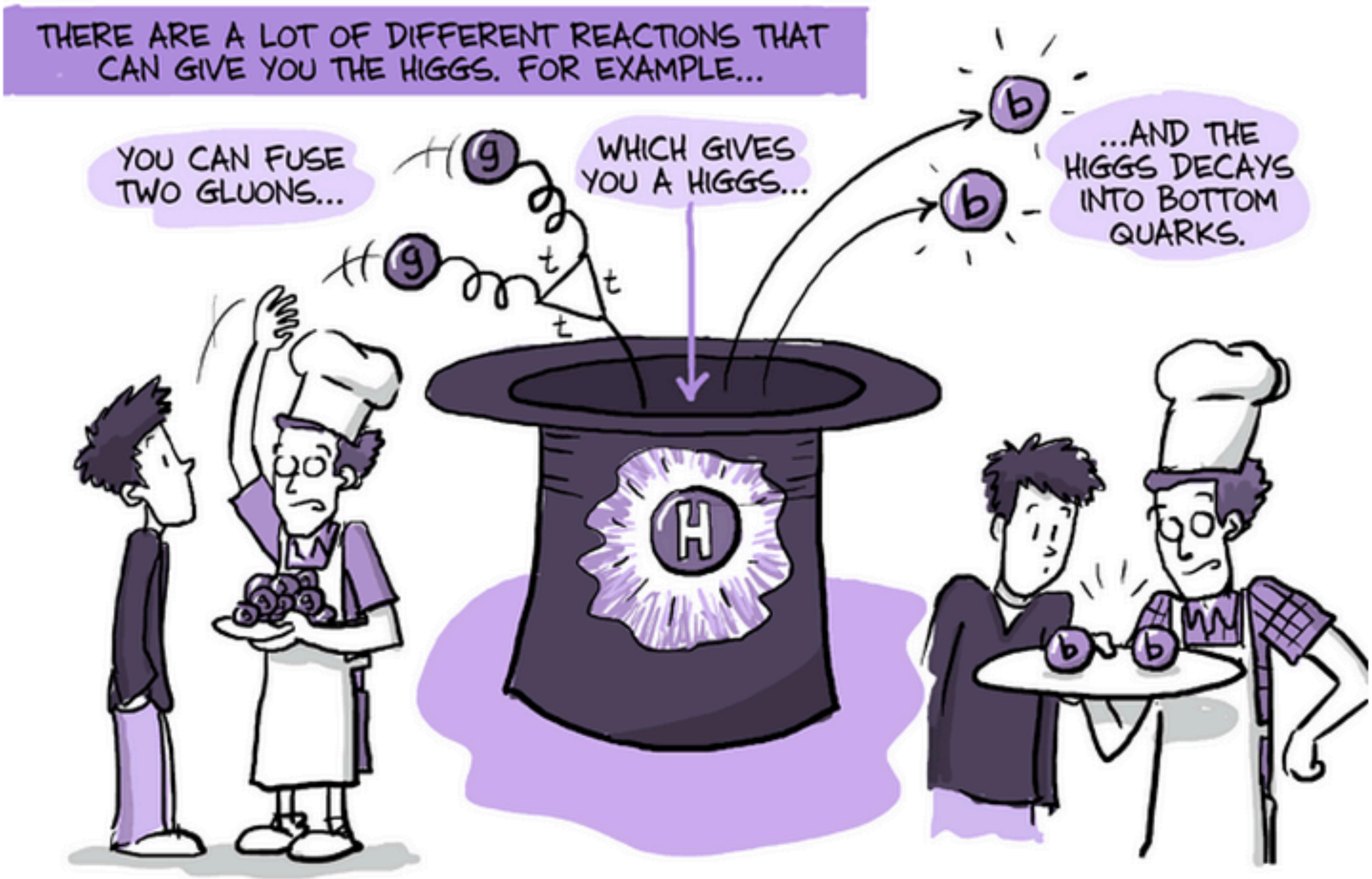
Traditional role of ML

Why do we need machine learning?

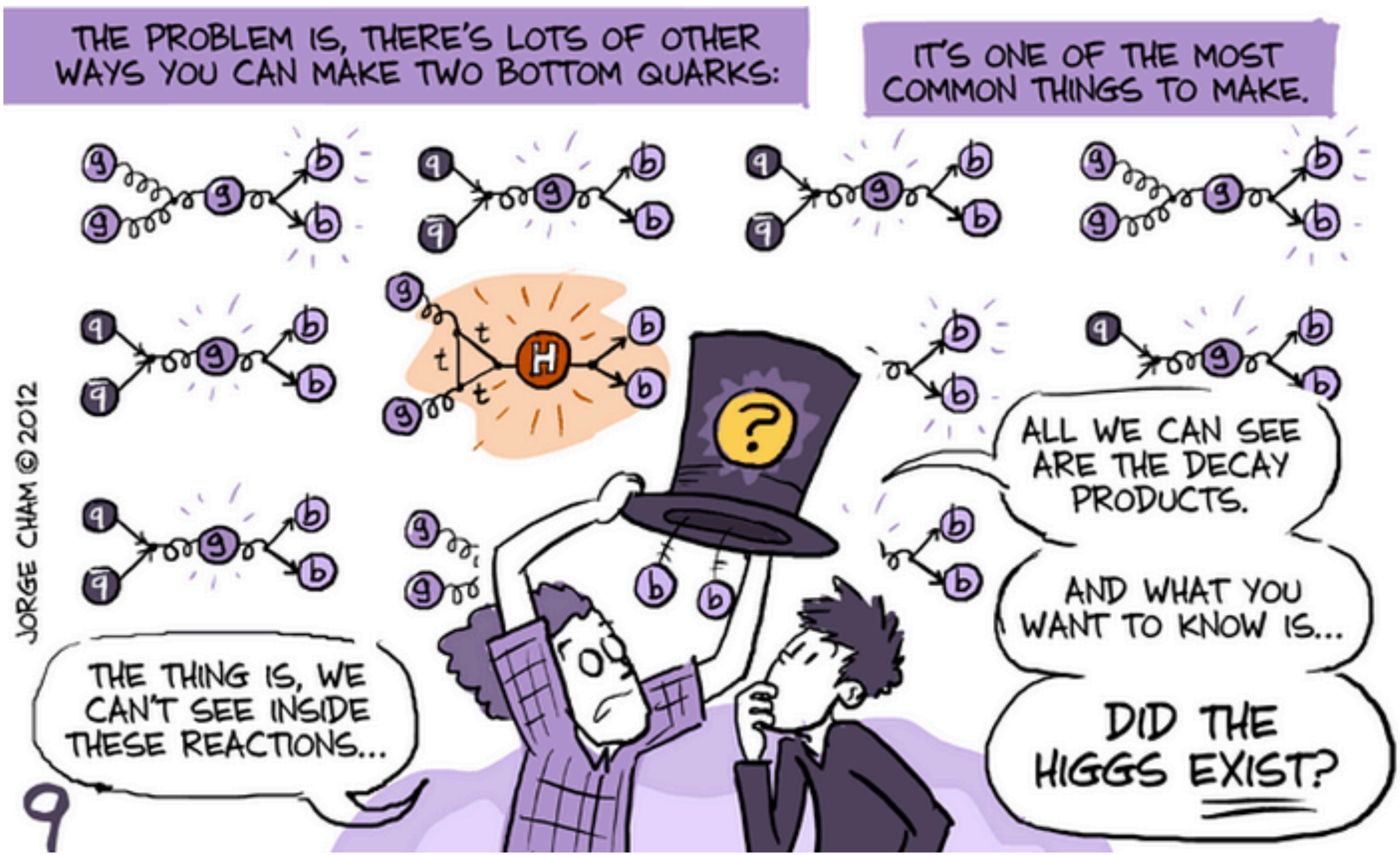
Traditional role of ML

Why do we need machine learning?

Making a new particle

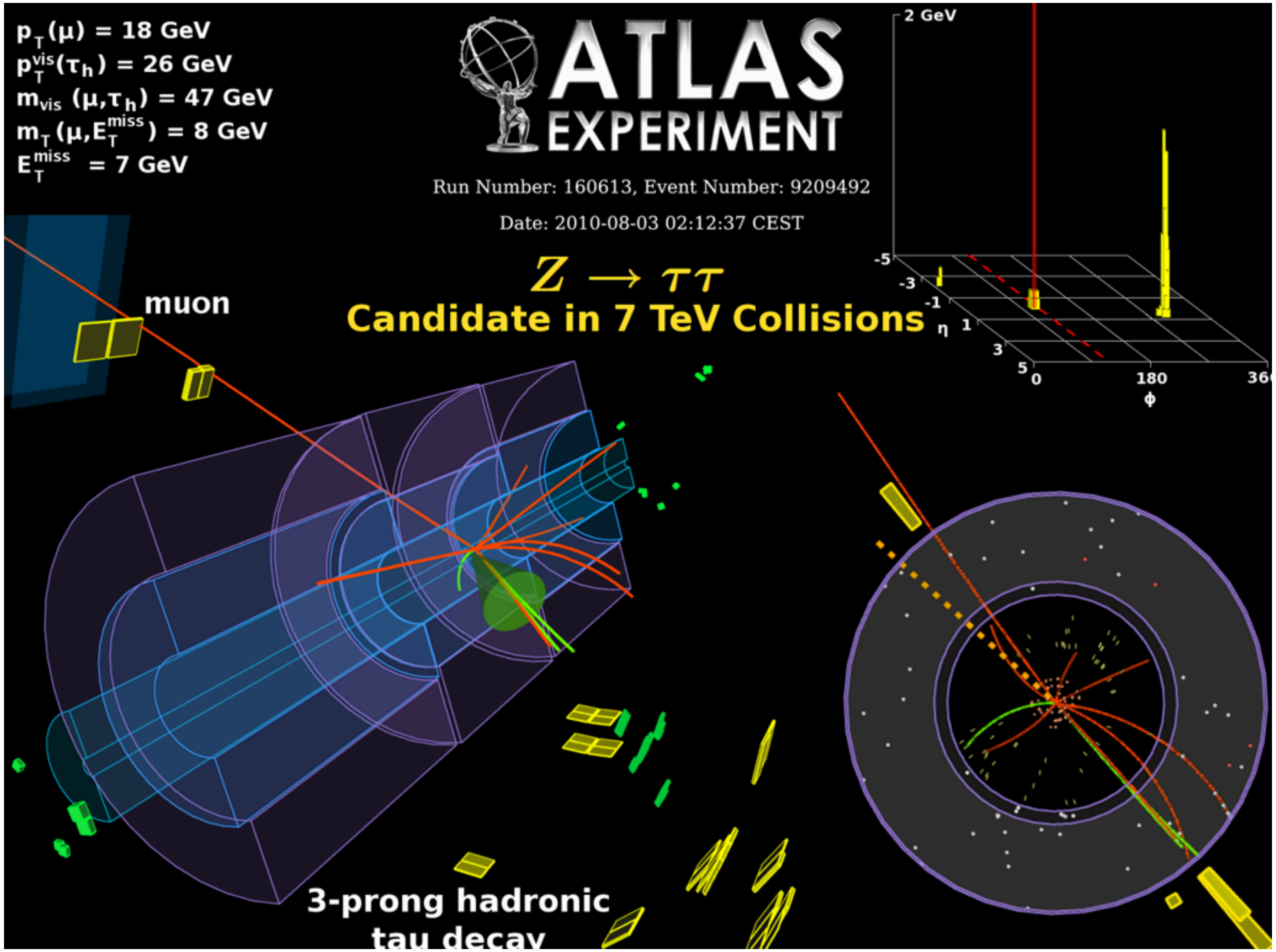


Backgrounds



Why statistics?

No event can be unambiguously interpreted.

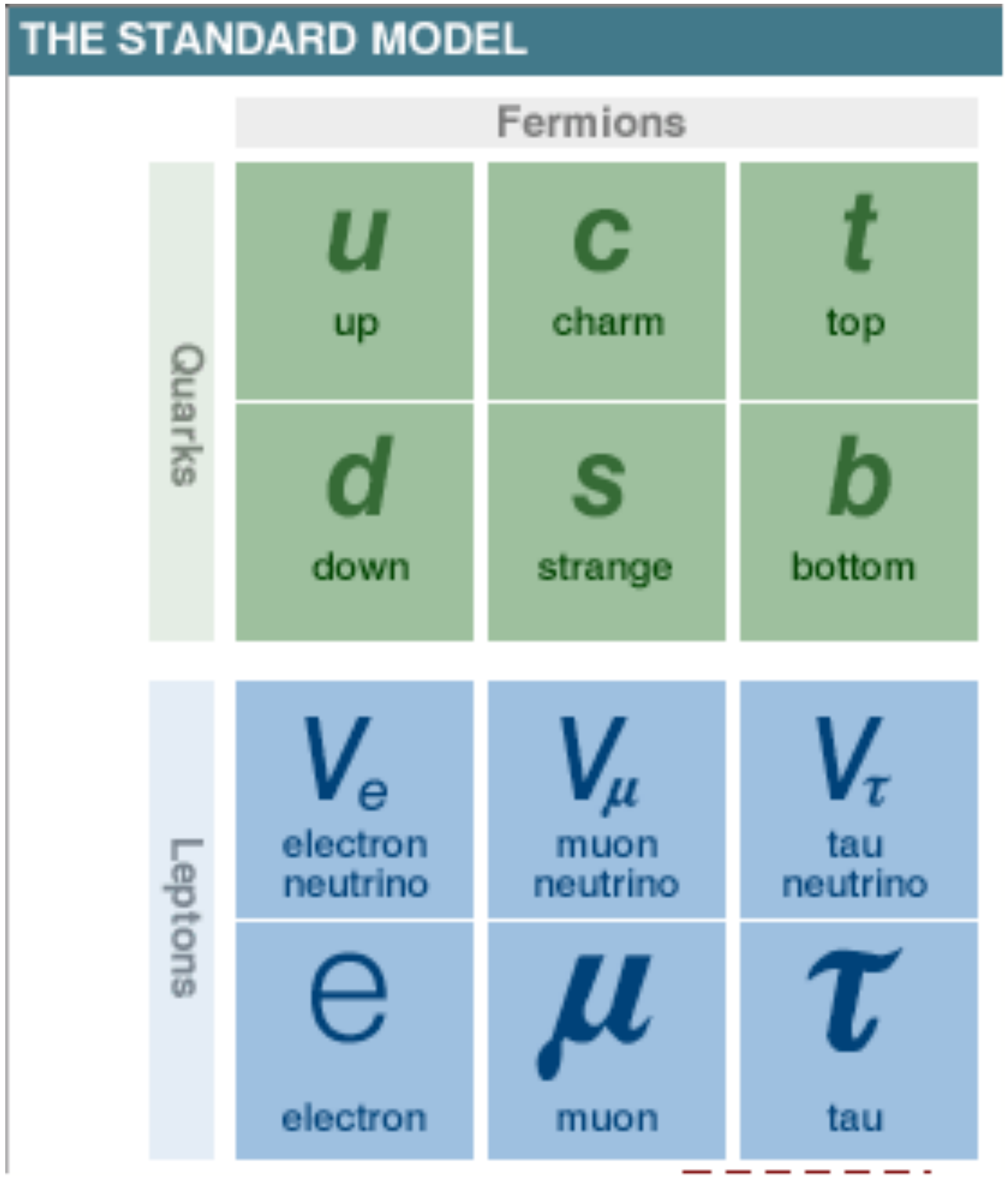


The nature of our data demands it.

Hypothesis testing

To search for a new particle, we compare the predictions of two hypotheses:

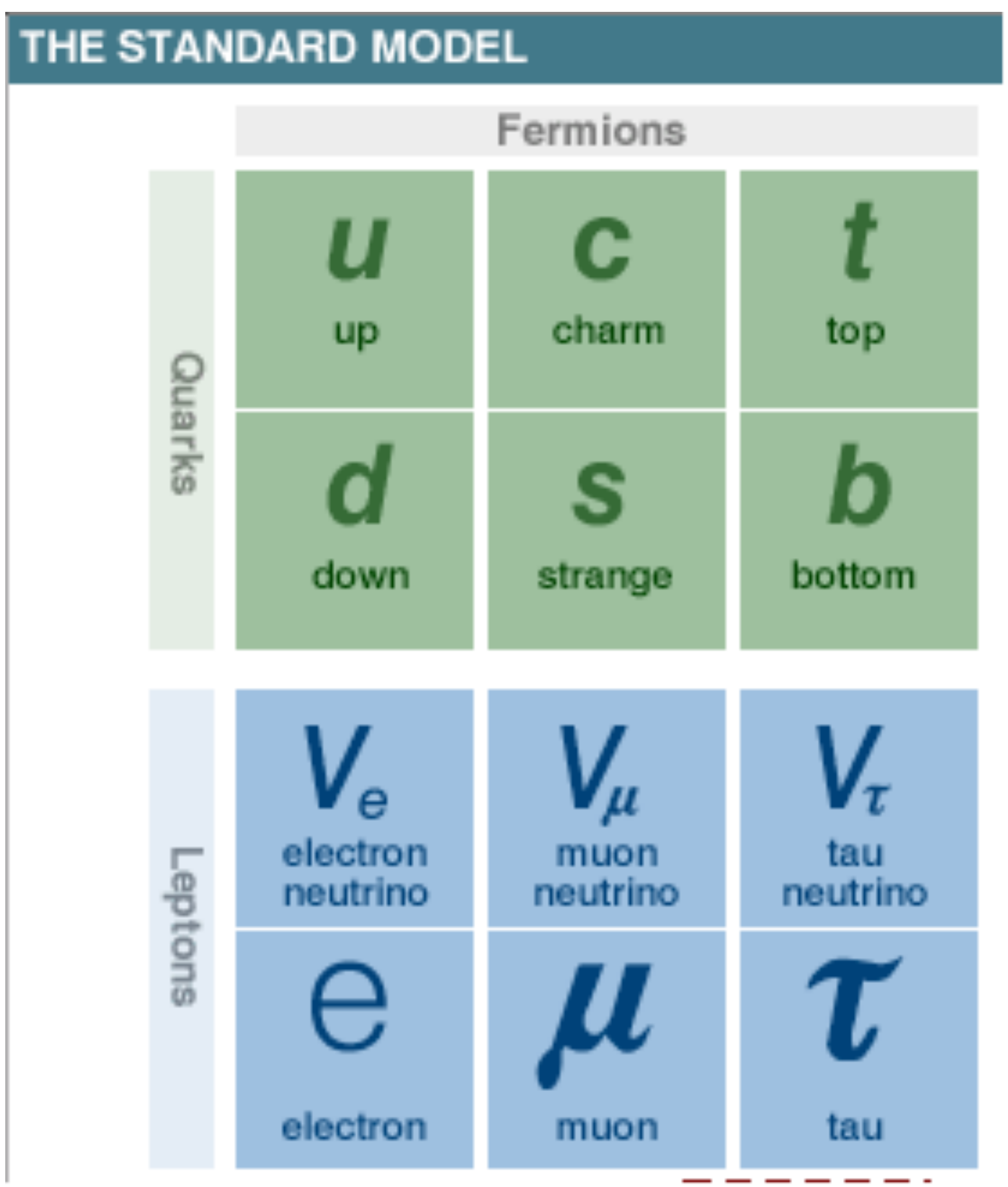
1\.



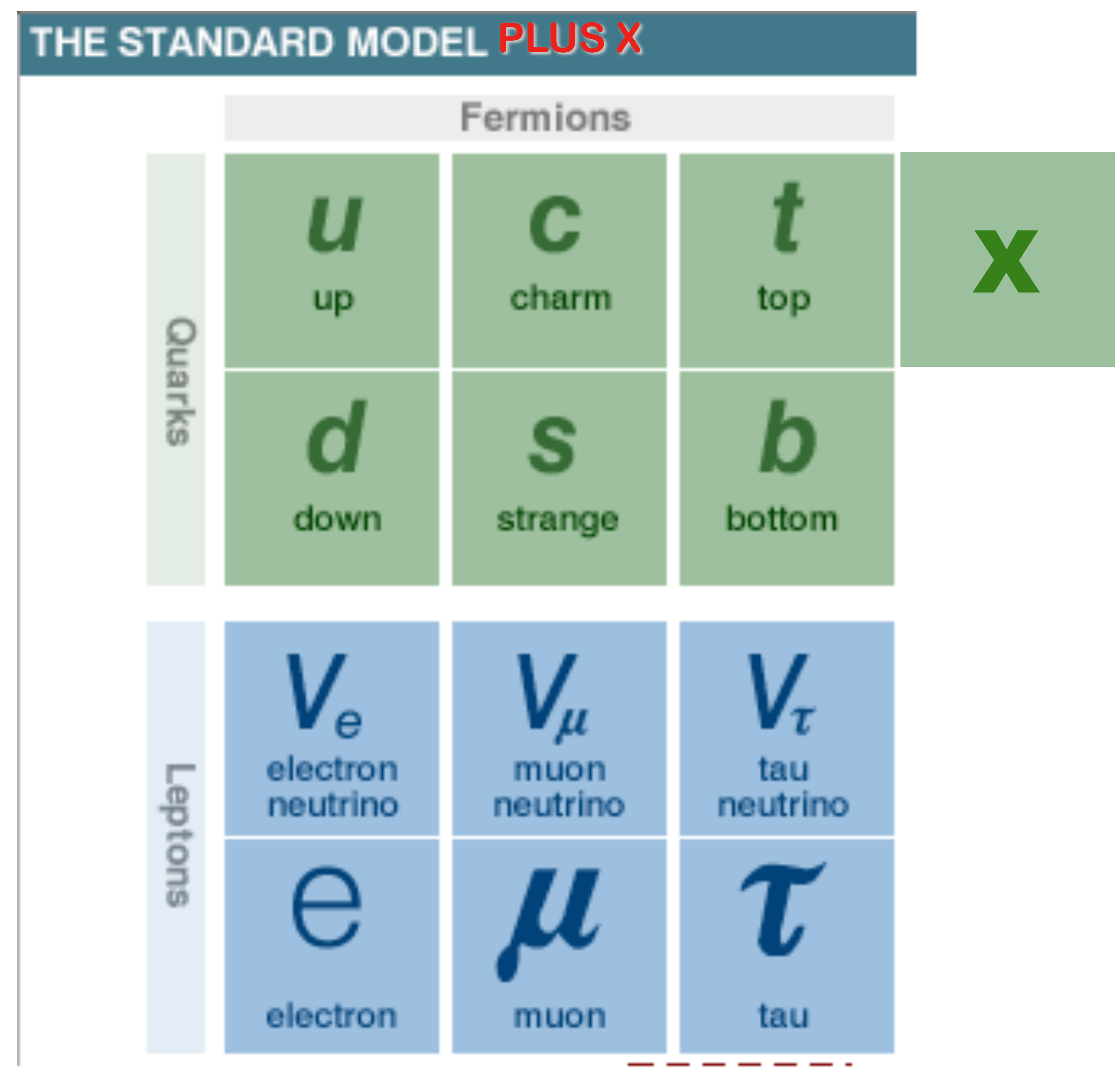
Hypothesis testing

To search for a new particle, we compare the predictions of two hypotheses:

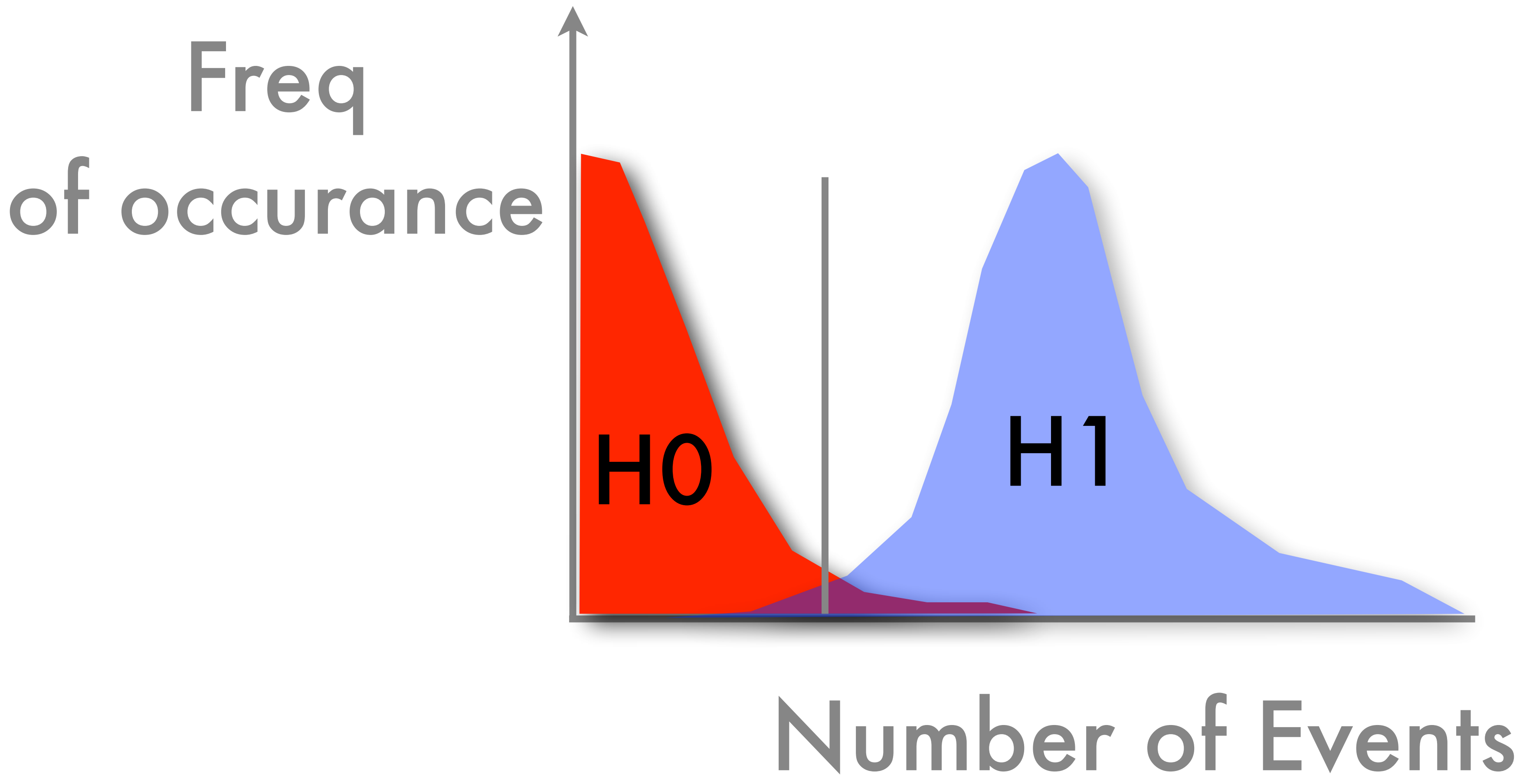
1.



2.

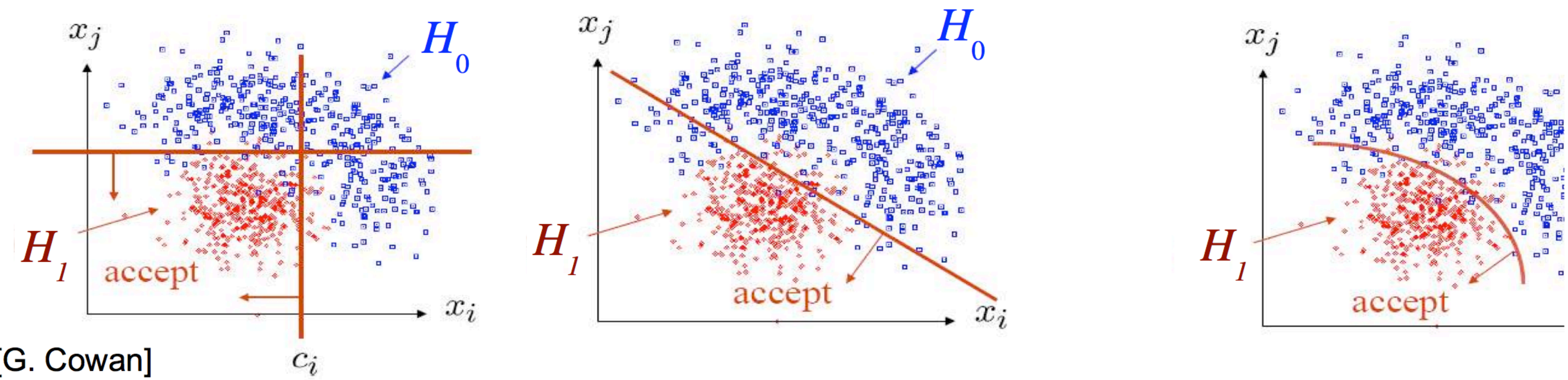


Example



A threshold makes sense.
Choice of position balances
false vs missed discovery

More complicated



[G. Cowan]

Neyman-Pearson

NP lemma says that the best statistic is the likelihood ratio:

$$\frac{P(x|H_1)}{P(x|H_0)} > k_\alpha$$

data

theory

(Gives smallest missed discovery rate
for fixed false discovery rate)

Functional space

No problem

If you can calculate:

$$\frac{P(x|H_1)}{P(x|H_0)} > k_\alpha$$

For which you need:

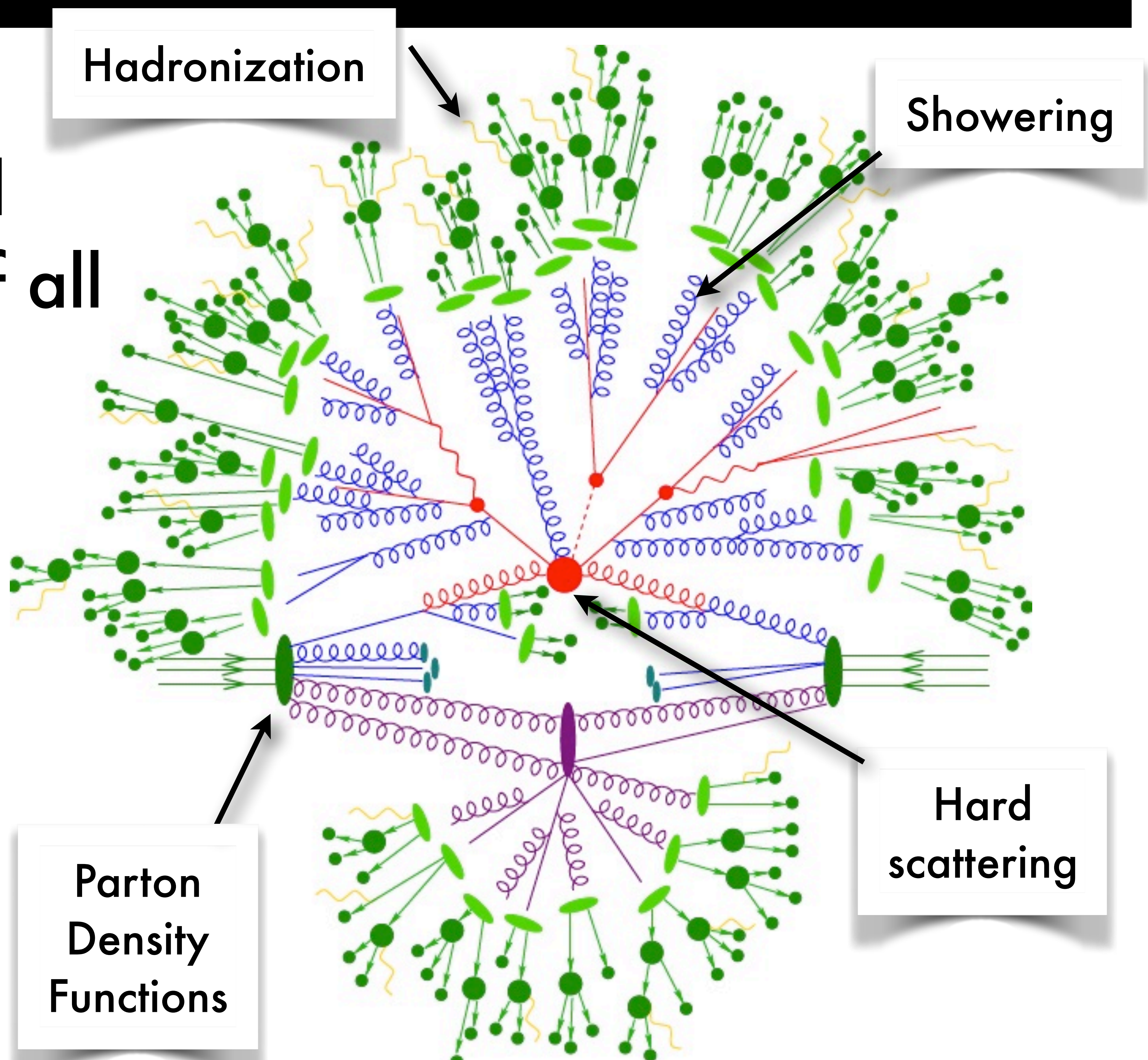
$$P(data|theory)$$

In general

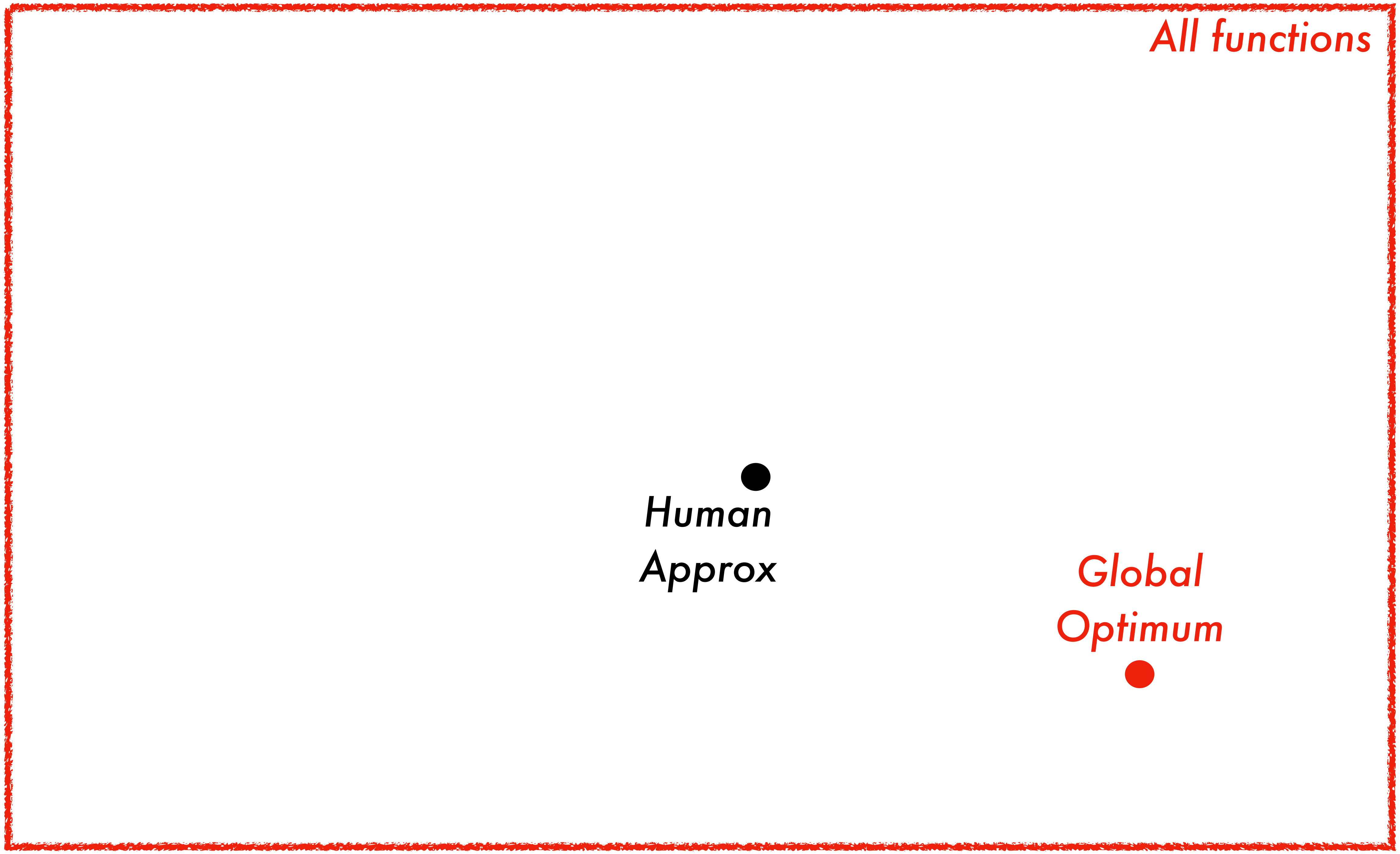
We have a good understanding of all of the pieces

Do we have

$P(\textcolor{blue}{data} \,|\, \textcolor{red}{theory})$



Functional space



In general

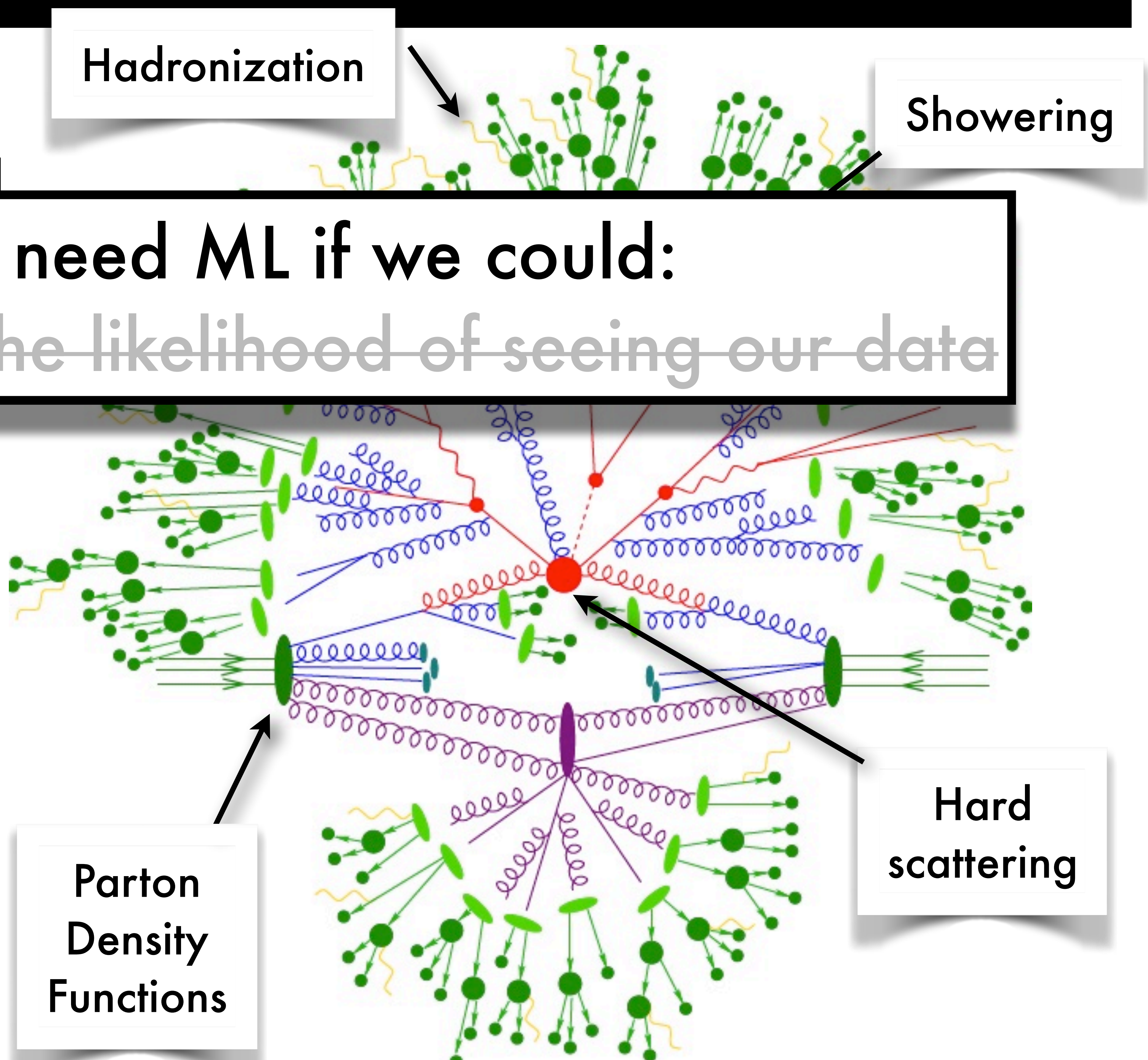
We have a good
und
of t

We wouldn't need ML if we could:

- ~~Express the likelihood of seeing our data~~

Do we have

$P(\color{blue}{data} \mid \color{red}{theory})$?

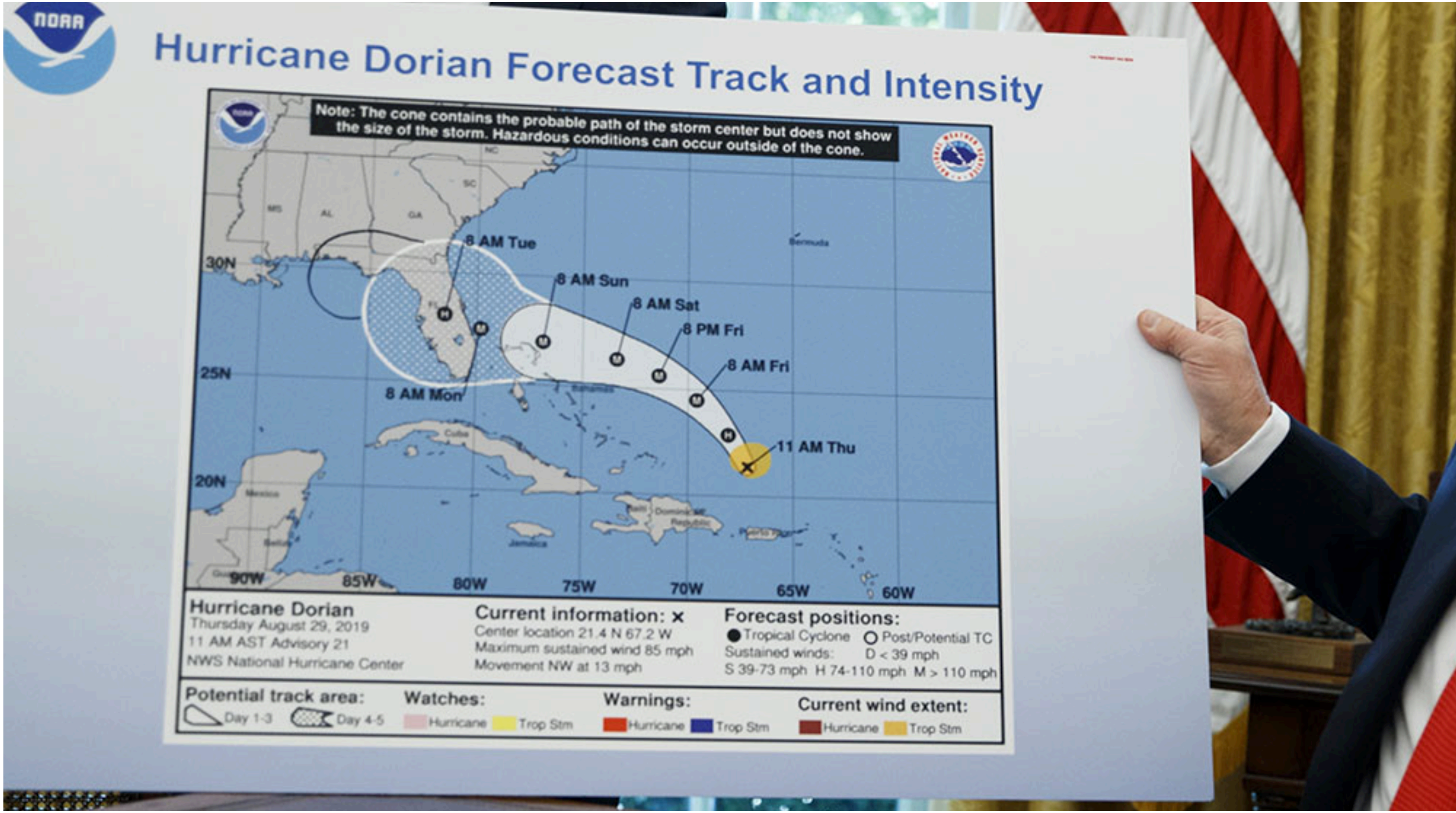


Darn

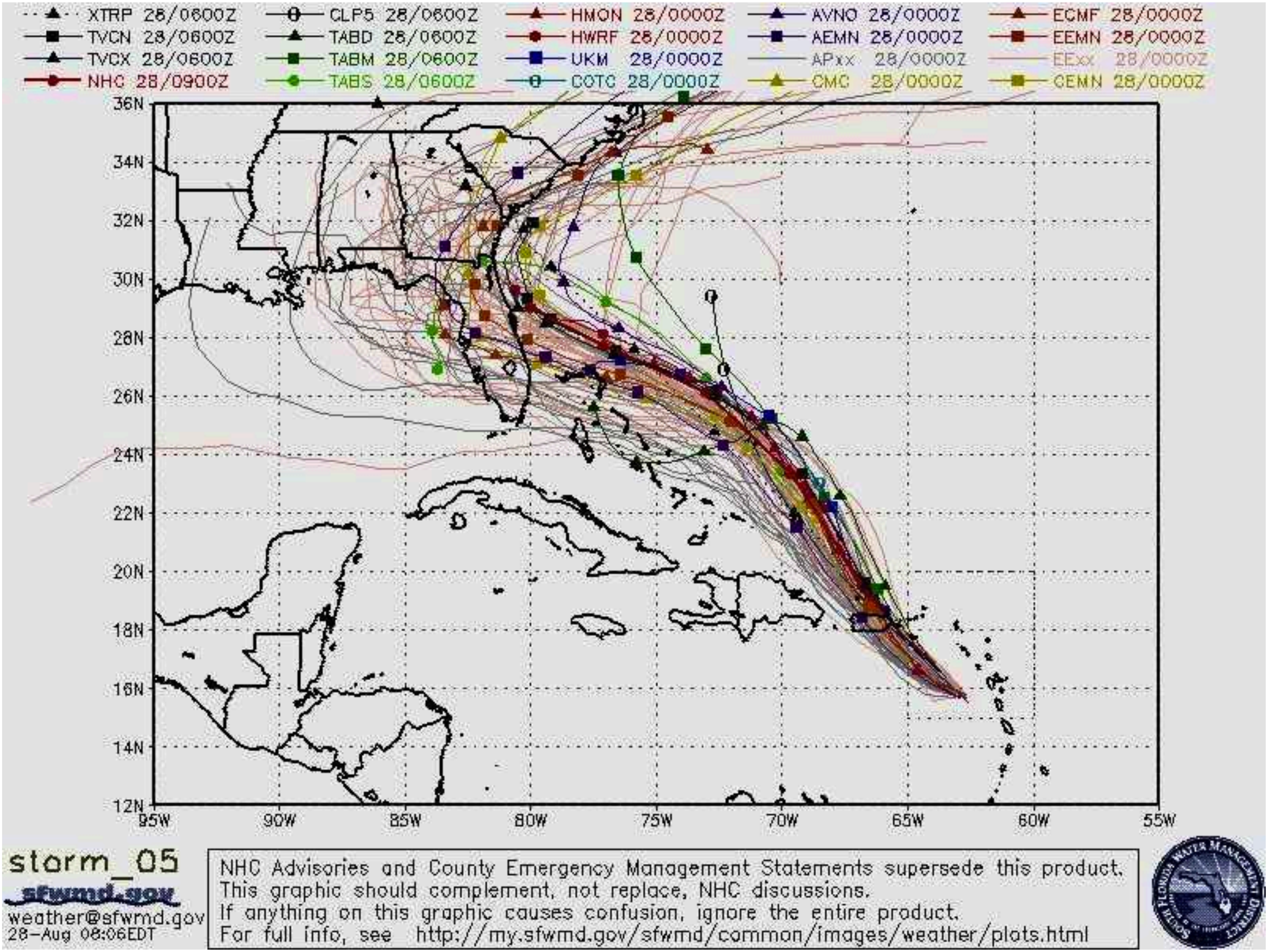
We can't calculate

$$P(data|theory)$$

.... but we can simulate it!



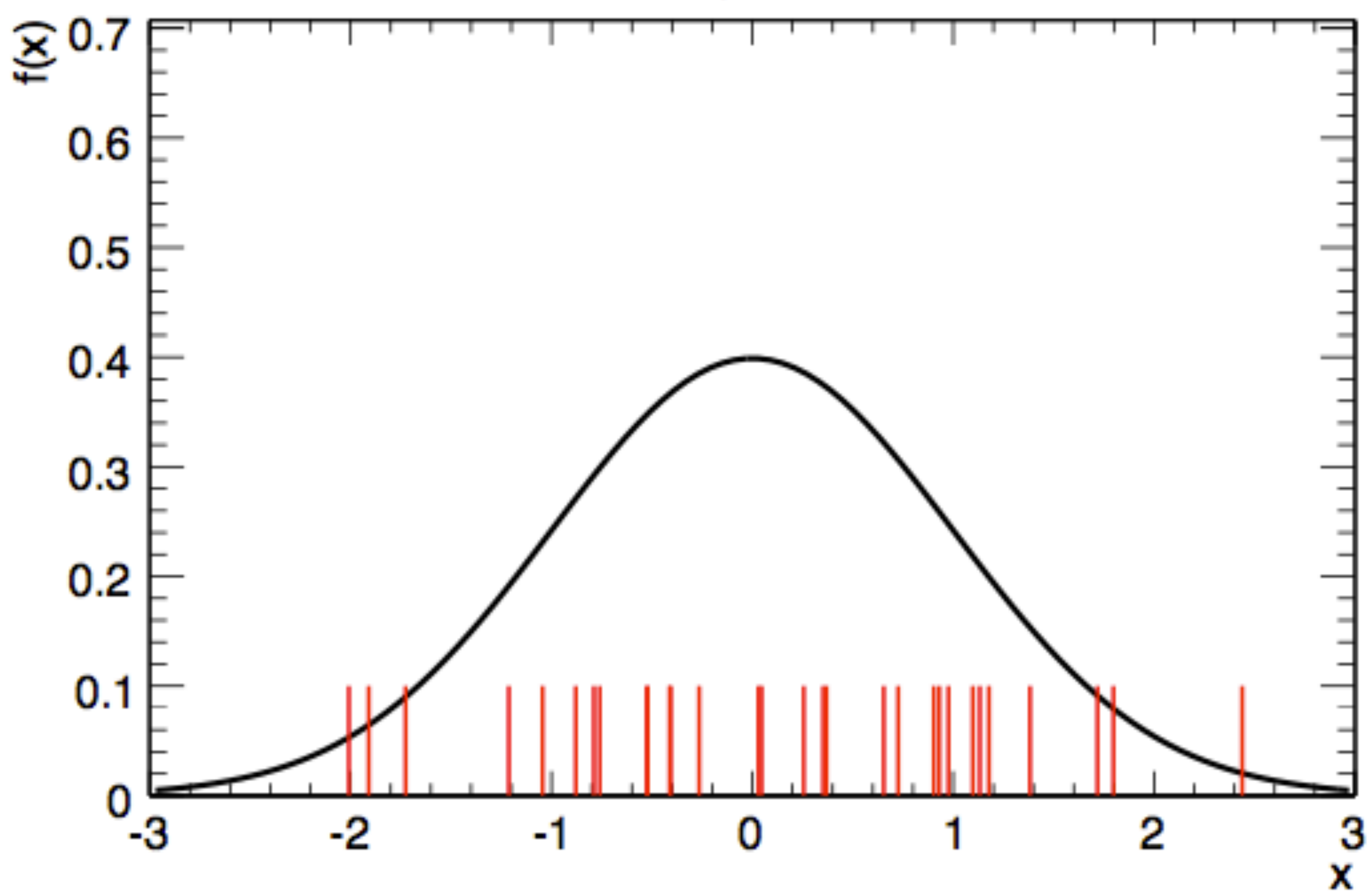
NOAA
Hurricane Dorian Forecast Track and Intensity
Note: The cone contains the probable path of the storm center but does not show the size of the storm. Hazardous conditions can occur outside of the cone.
8 AM Tue
8 AM Sun
8 AM Sat
8 PM Fri
8 AM Fri
8 AM Mon
11 AM Thu
30N
25N
20N
90W
85W
80W
75W
70W
65W
60W
Hurricane Dorian
Thursday August 29, 2019
11 AM AST Advisory 21
NWS National Hurricane Center
Current information: x
Center location 21.4 N 67.2 W
Maximum sustained wind 85 mph
Movement NW at 13 mph
Forecast positions:
Tropical Cyclone
Post/Potential TC
Sustained winds:
D < 39 mph
S 39-73 mph H 74-110 mph M > 110 mph
Potential track area:
Day 1-3
Day 4-5
Watches:
Hurricane
Trop Stm
Warnings:
Hurricane
Trop Stm
Current wind extent:
Hurricane
Trop Stm


XTRP 28/0600Z
TVCN 28/0600Z
TVCX 28/0600Z
NHC 28/0900Z
CLP5 28/0600Z
TABD 28/0600Z
TABM 28/0600Z
TABS 28/0600Z
HMON 28/0000Z
HWRF 28/0000Z
UKM 28/0000Z
COTC 28/0000Z
AVNO 28/0000Z
AEMN 28/0000Z
APxx 28/0000Z
CMC 28/0000Z
ECMF 28/0000Z
EEMN 28/0000Z
EExx 28/0000Z
CEMN 28/0000Z
36N
34N
32N
30N
28N
26N
24N
22N
20N
18N
16N
14N
12N
95W
90W
85W
80W
75W
70W
65W
60W
55W
storm_05
sfwmd.gov
weather@sfwmd.gov
28-Aug 08:06EDT
NHC Advisories and County Emergency Management Statements supersede this product.
This graphic should complement, not replace, NHC discussions.
If anything on this graphic causes confusion, ignore the entire product.
For full info, see http://my.sfwmd.gov/sfwmd/common/images/weather/plots.html

The problem

Don't know PDF, have events drawn from PDF

$$f_{emp} = \frac{1}{N} \sum_{i}^{N} \delta(x - x_i)$$



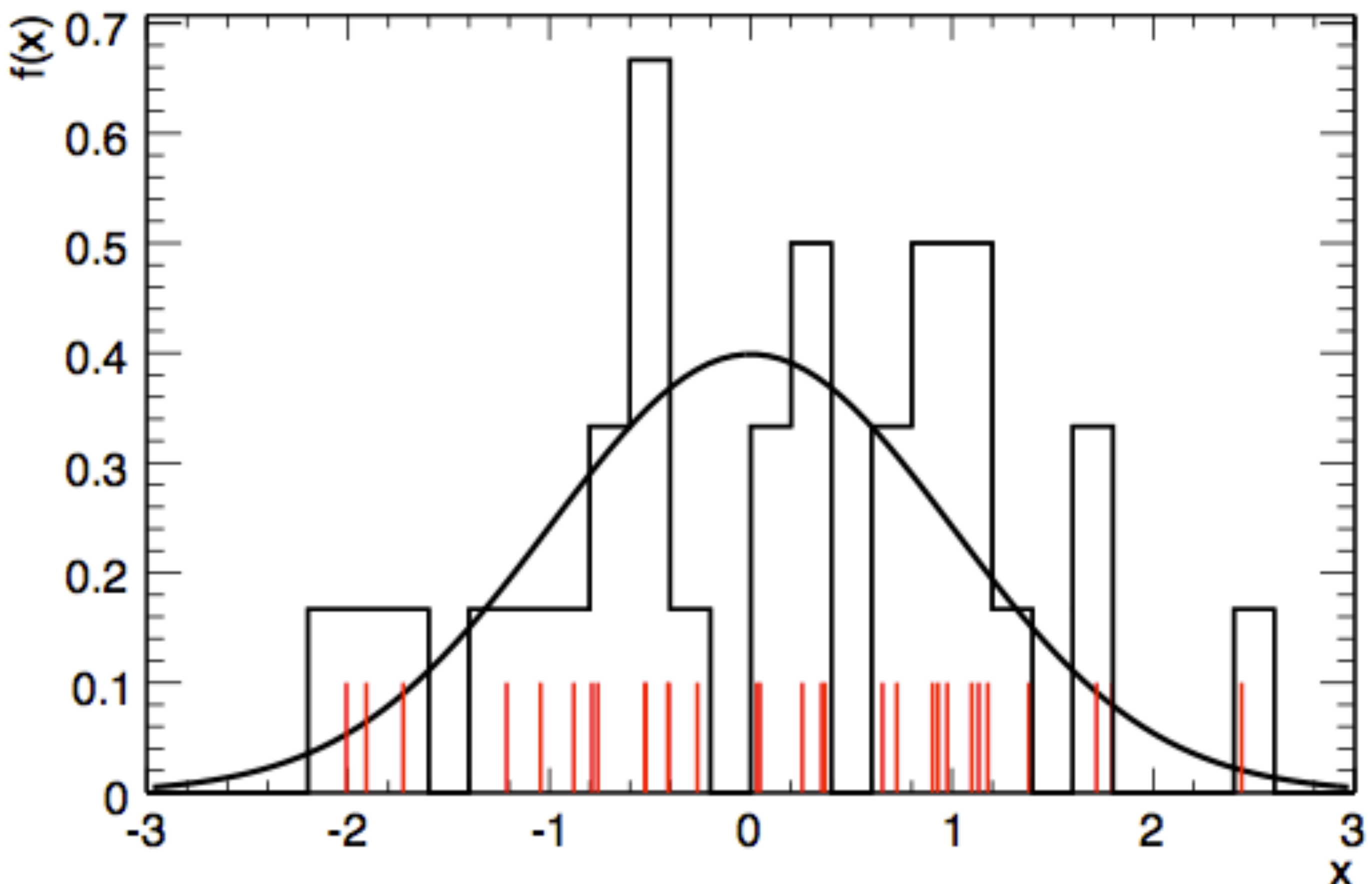
(K. Cranmer)

Need to recreate PDF

MC events to PDF

Simple approach : histogram

$$f^{w,s}_{hist}(x) = \frac{1}{N}\sum_i h_i^{w,s}$$



(K. Cranmer)

Curse of Dimensionality

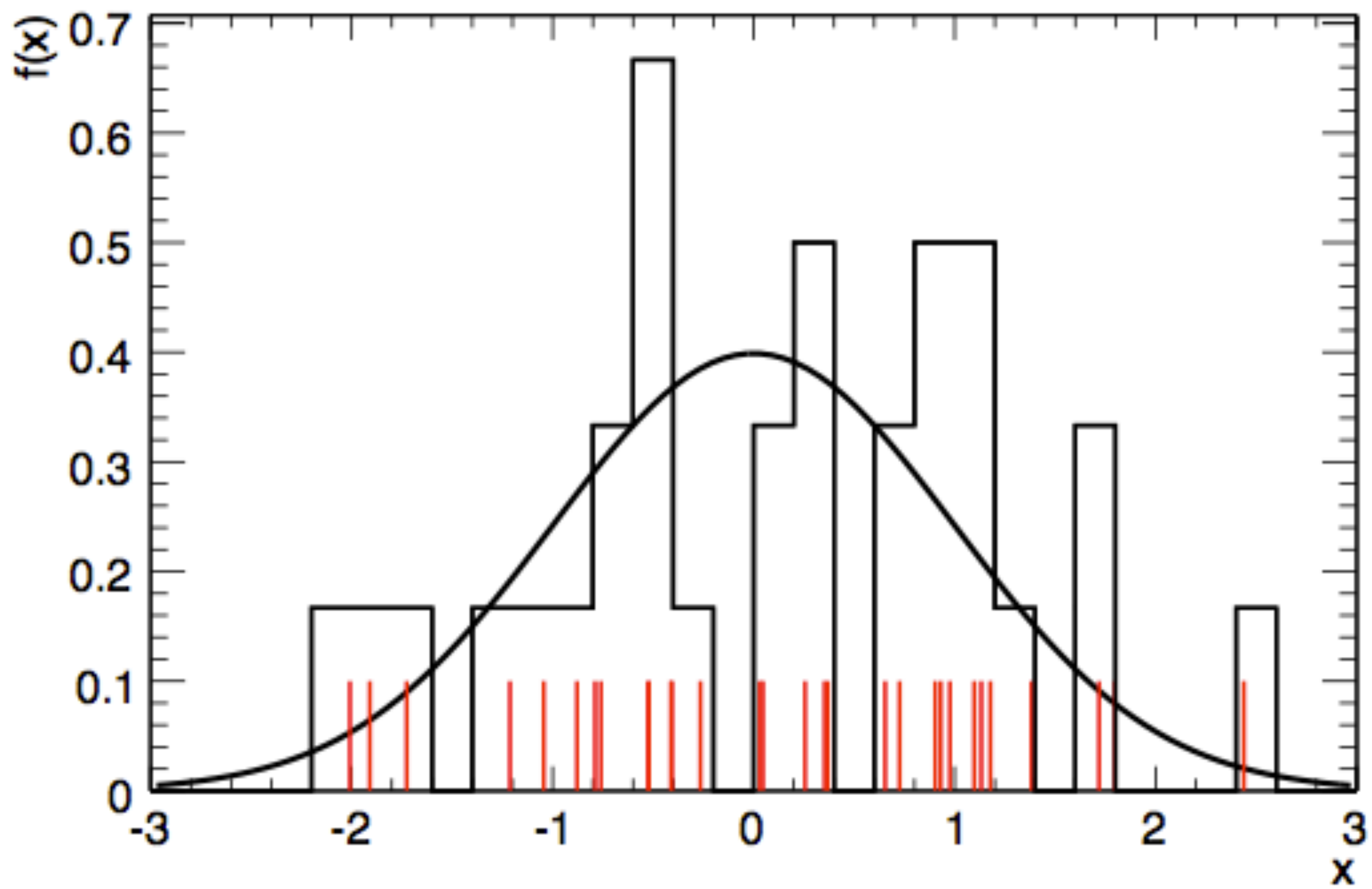
How many events
do you need
to describe a 1D
distribution? O(100)

An n-D distribution?

$O(100^n)$

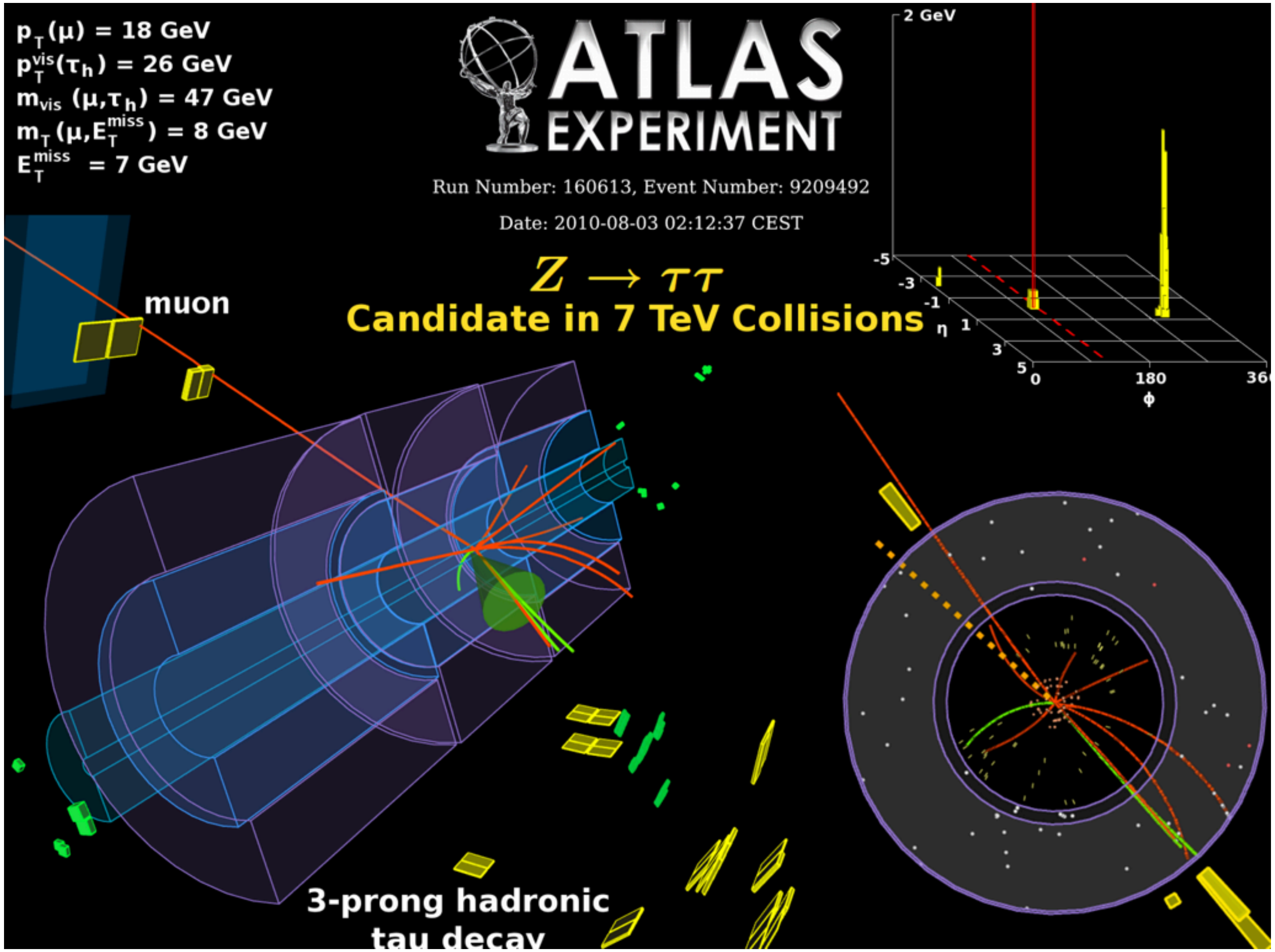
!!

$$f^{w,s}_{hist}(x) = \frac{1}{N}\sum_i h_i^{w,s}$$



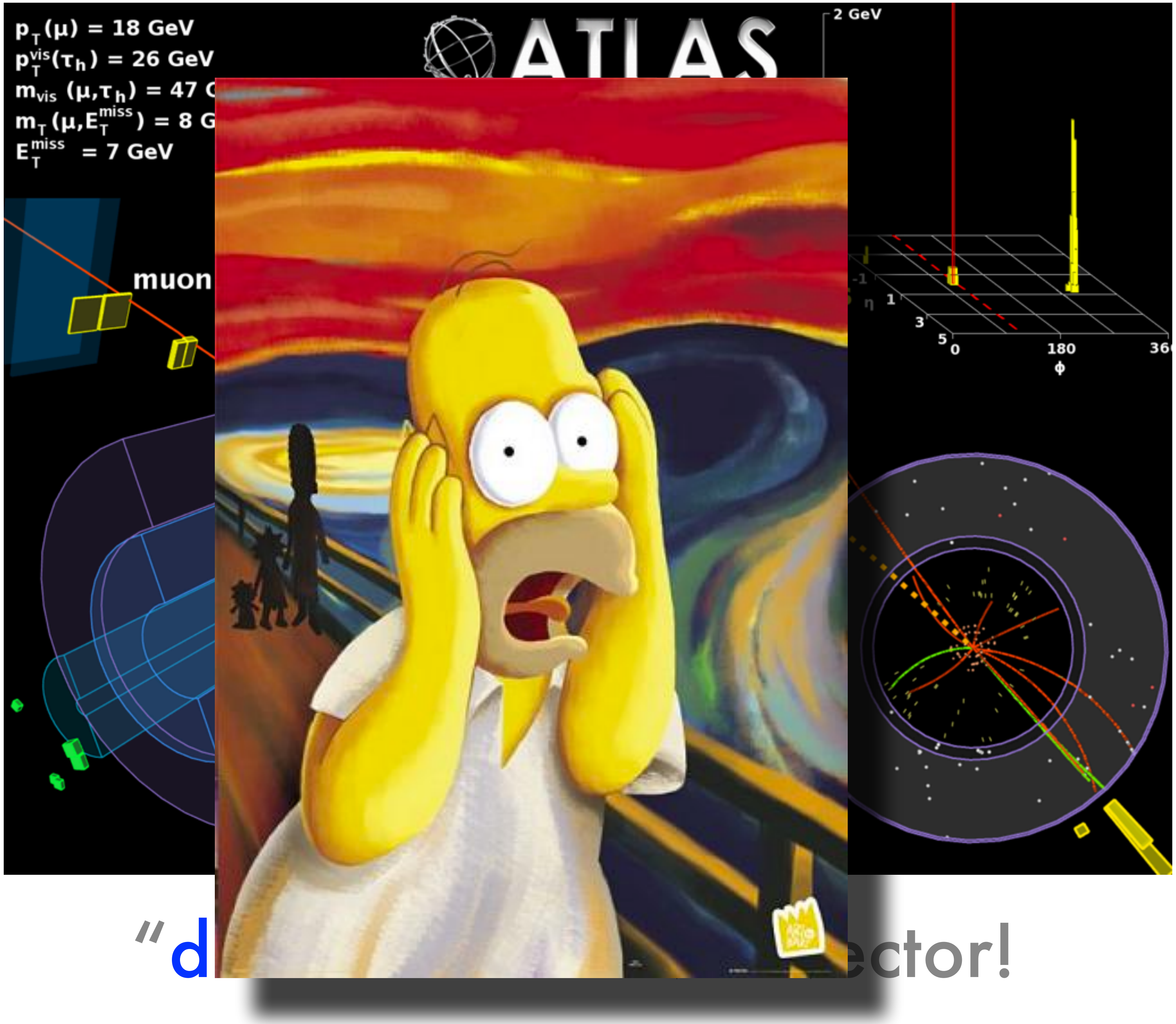
(K. Cranmer)

The nightmare



"data" is a 100M-d vector!

The nightmare



“d ector!

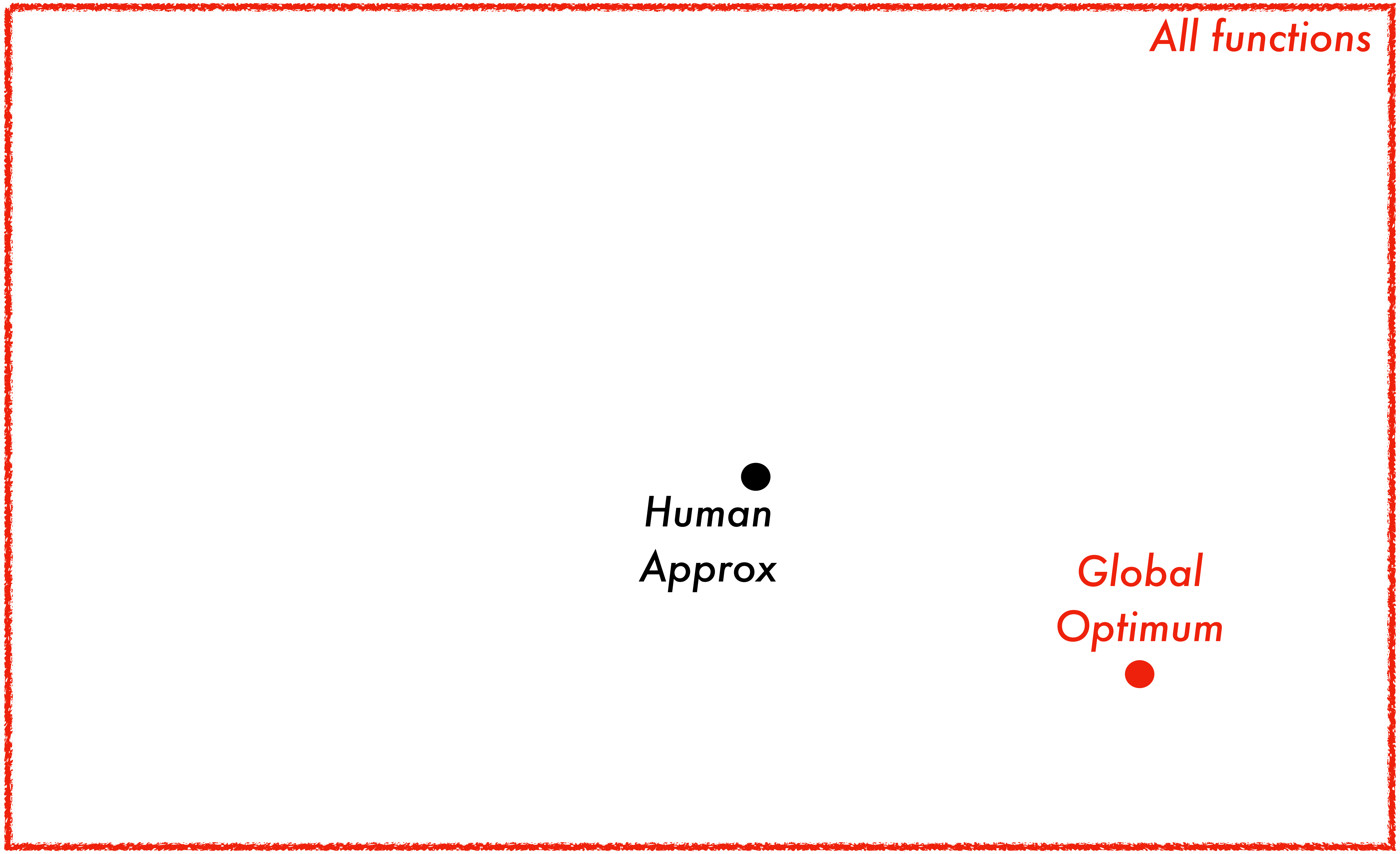
The nightmare

We wouldn’t need ML if we could:

- ~~Express the likelihood of seeing our data~~
- ~~Access infinite computing resources~~
- ~~Develop infinitely-fast simulation~~

“d ector!

Functional space



Task for ML

Find a function:

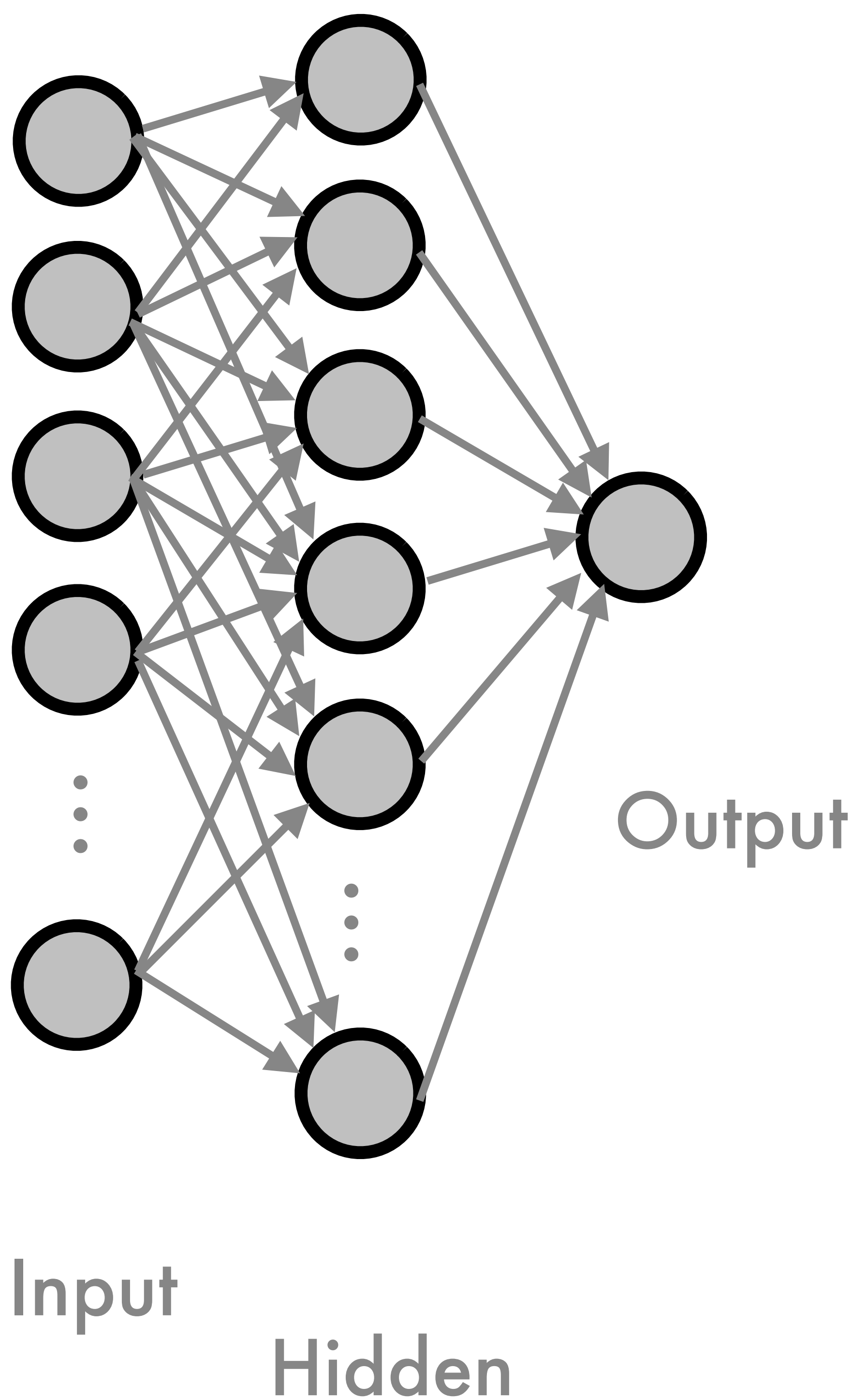
$$f(\bar{x}) : \mathbb{R}^N \rightarrow \mathbb{R}^1$$

which contains the same hypothesis testing power as

$$\frac{P(x|H_1)}{P(x|H_0)} > k_\alpha$$

How complex?

Essentially a functional fit with many parameters

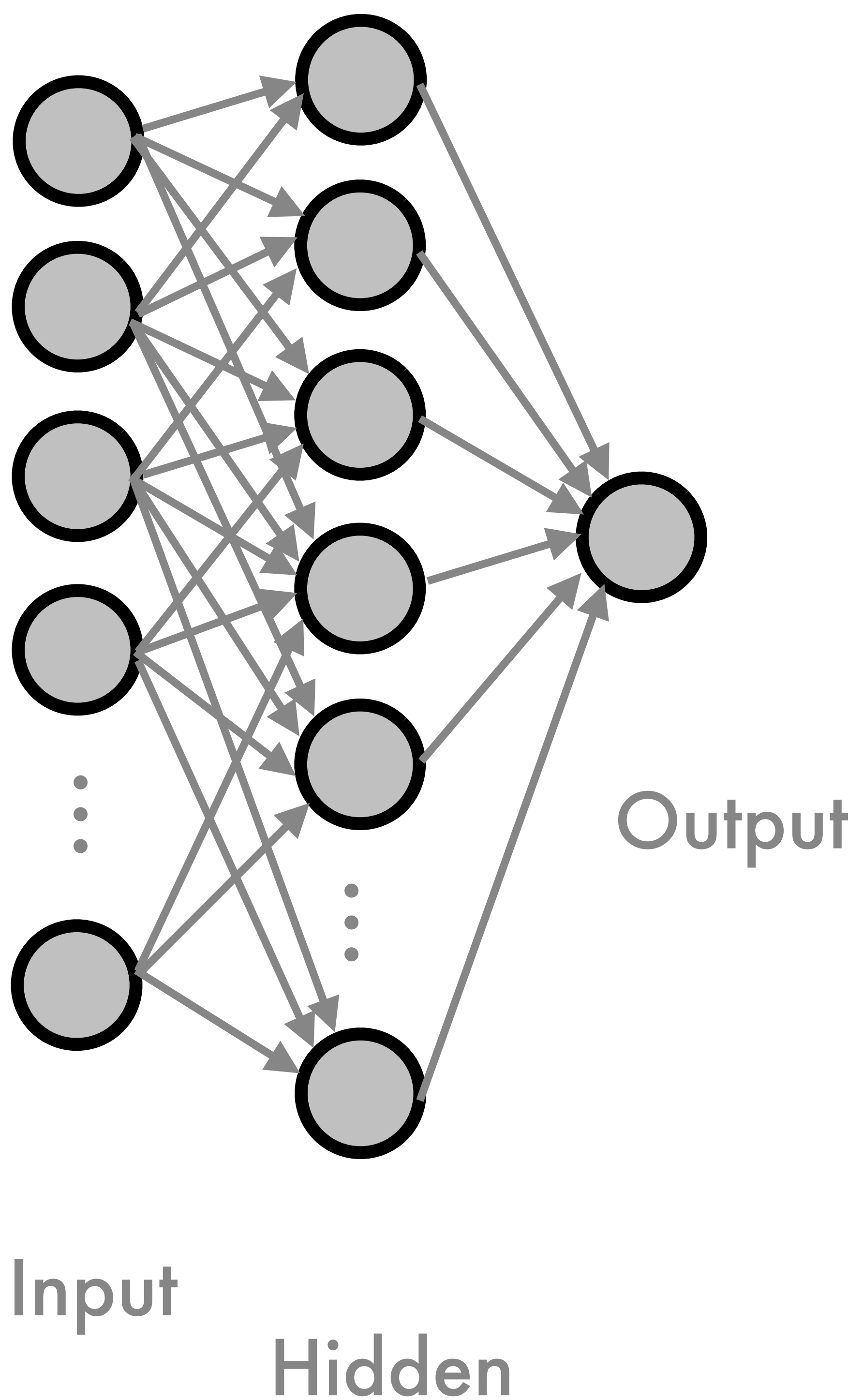


Single hidden layer

In theory any function
can be learned with
a single hidden layer.

How complex?

Essentially a functional fit with many parameters

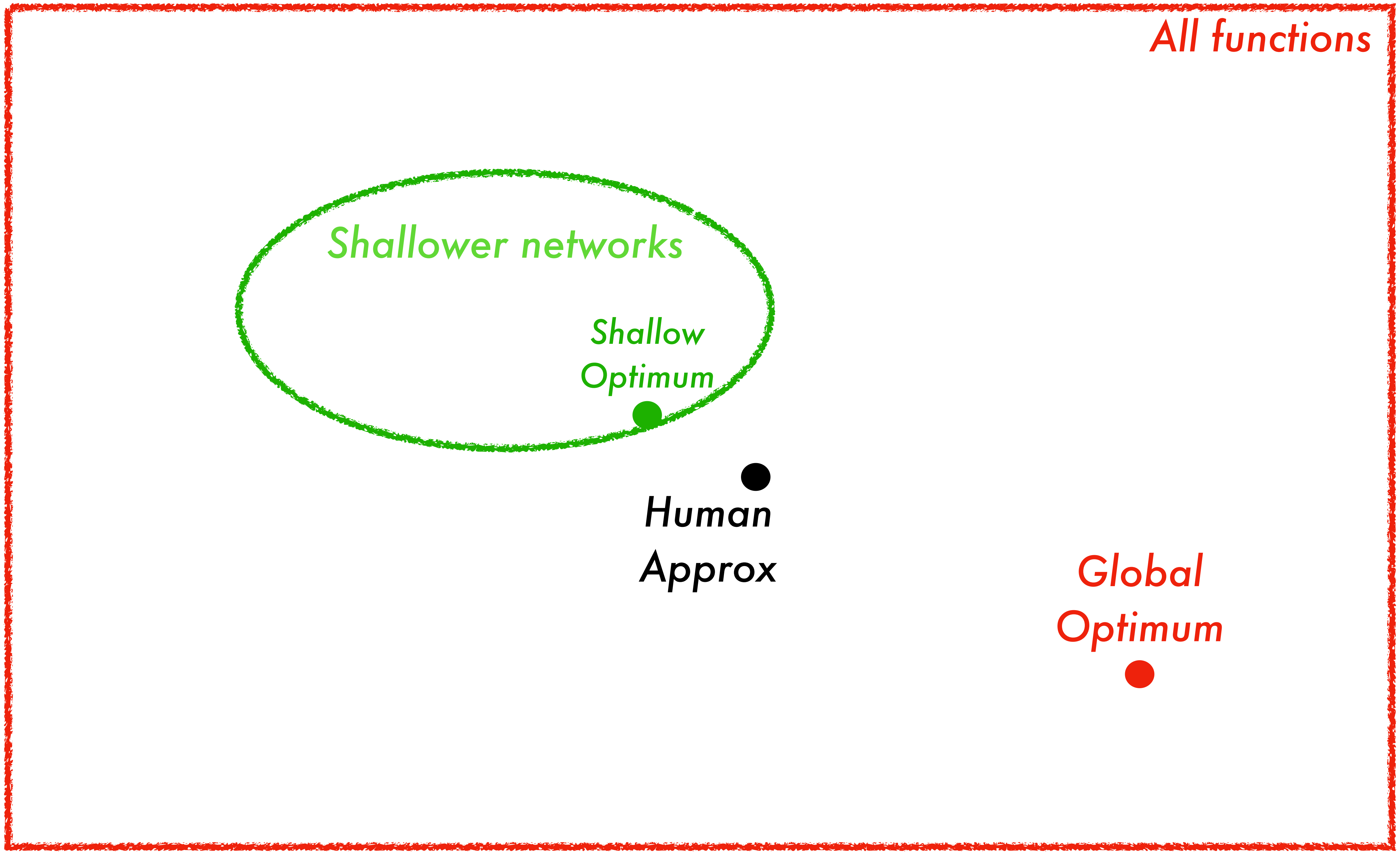


<u>Single hidden layer</u>

In theory any function can be learned with a single hidden layer.

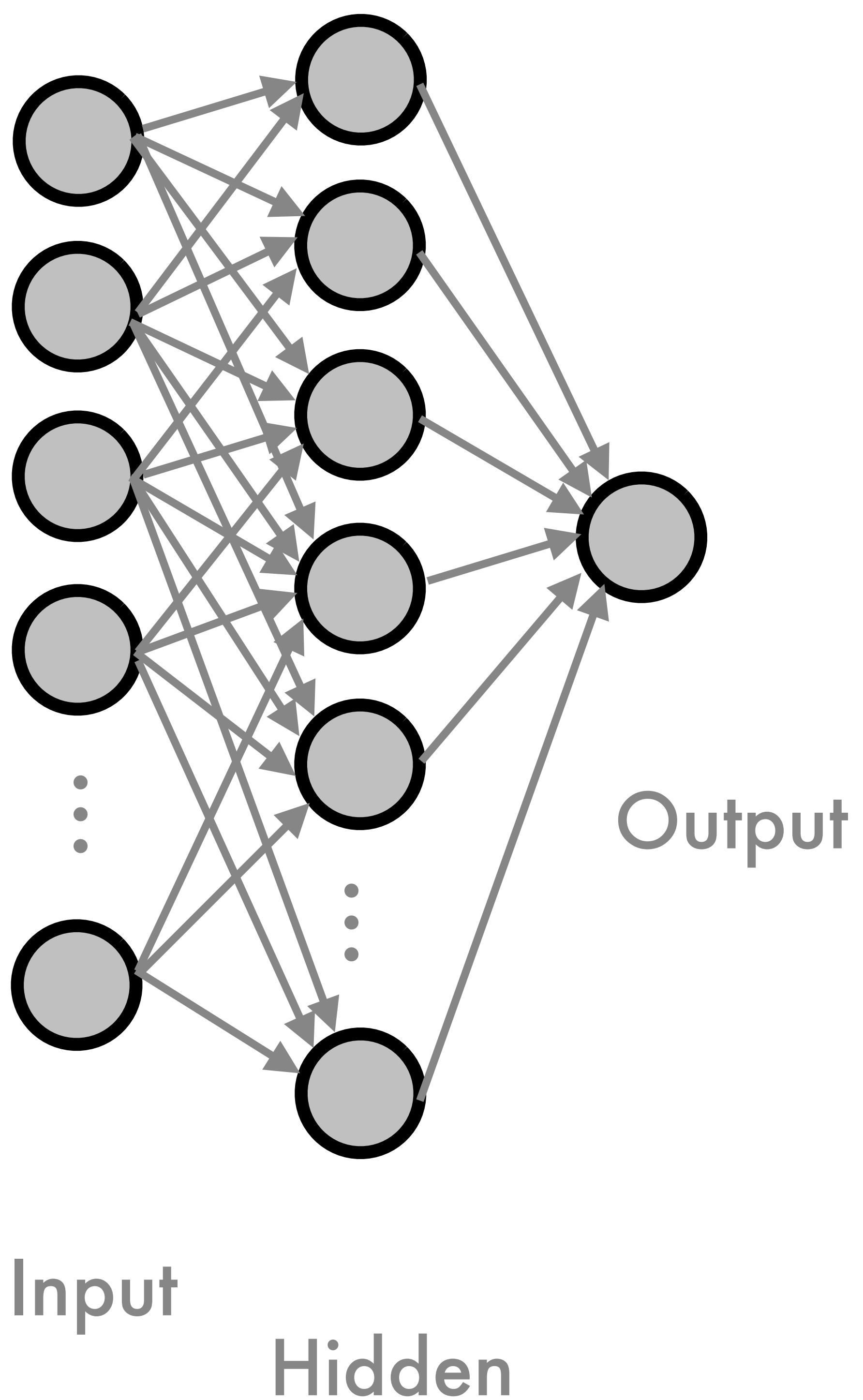
But might require very large hidden layer to learn non-linear functions

Shallow space



Neural Networks

Essentially a functional fit with many parameters



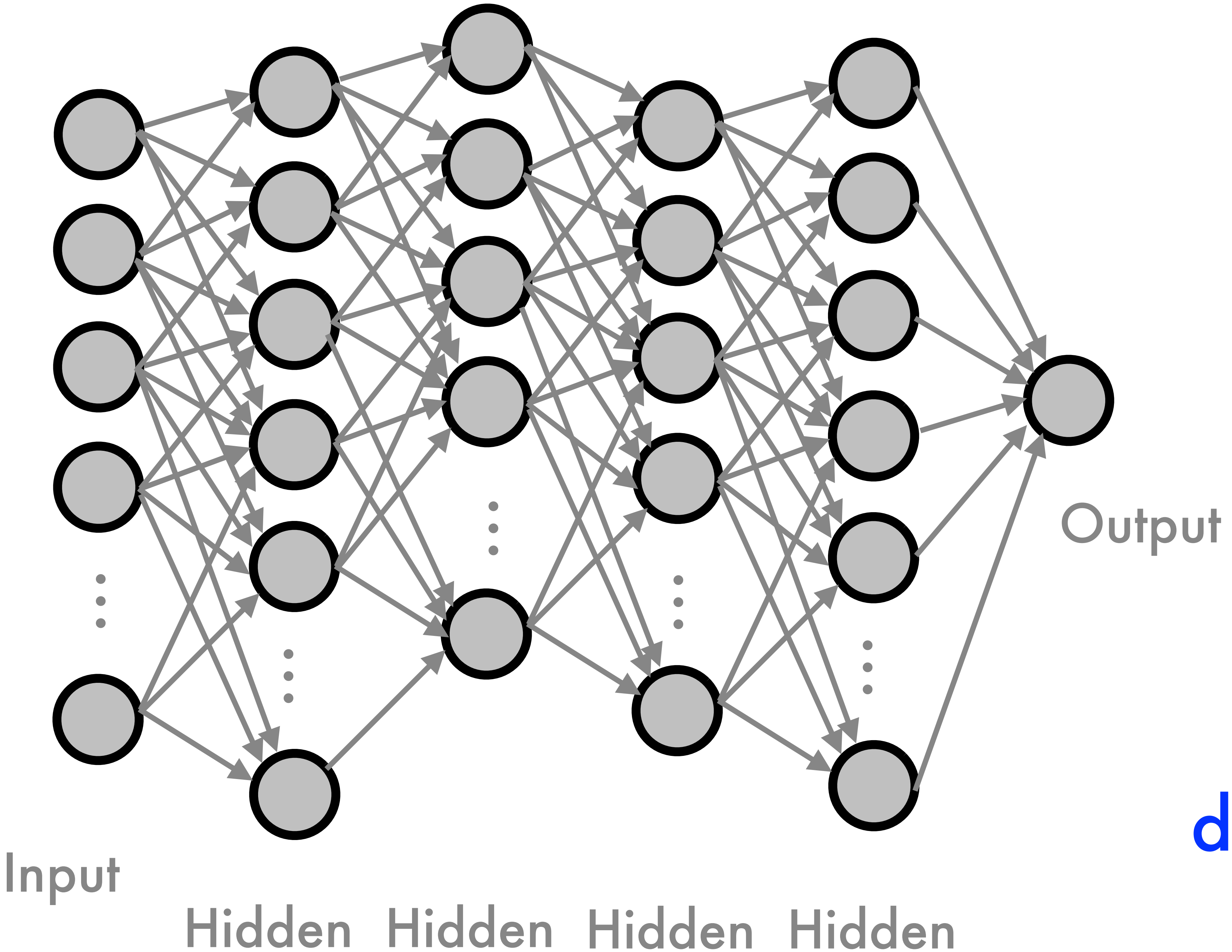
<u>Consequence:</u>

Networks are not good
at learning non-linear functions.
(like invariant masses!)

<u>In short:</u>

Couldn't just throw data at NN.

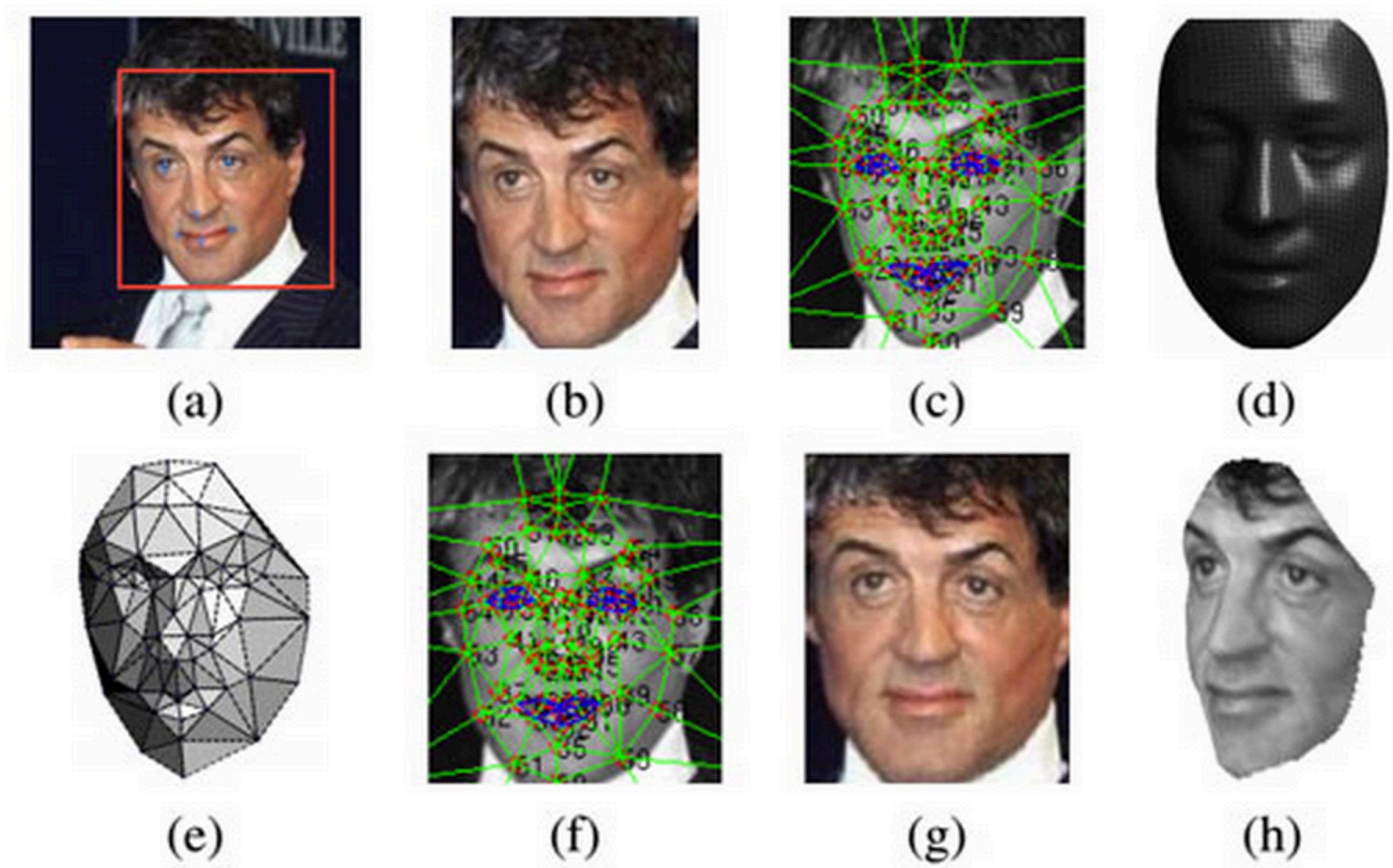
Deep networks



New tools let us train deep networks.

How well do they work?

Real world applications



Head turn: DeepFace uses a 3-D model to rotate faces, virtually, so that they face the camera. Image (a) shows the original image, and (g) shows the final, corrected version.

Paper

nature COMMUNICATIONS

ARTICLE

Received 19 Feb 2014 | Accepted 4 Jun 2014 | Published 2 Jul 2014

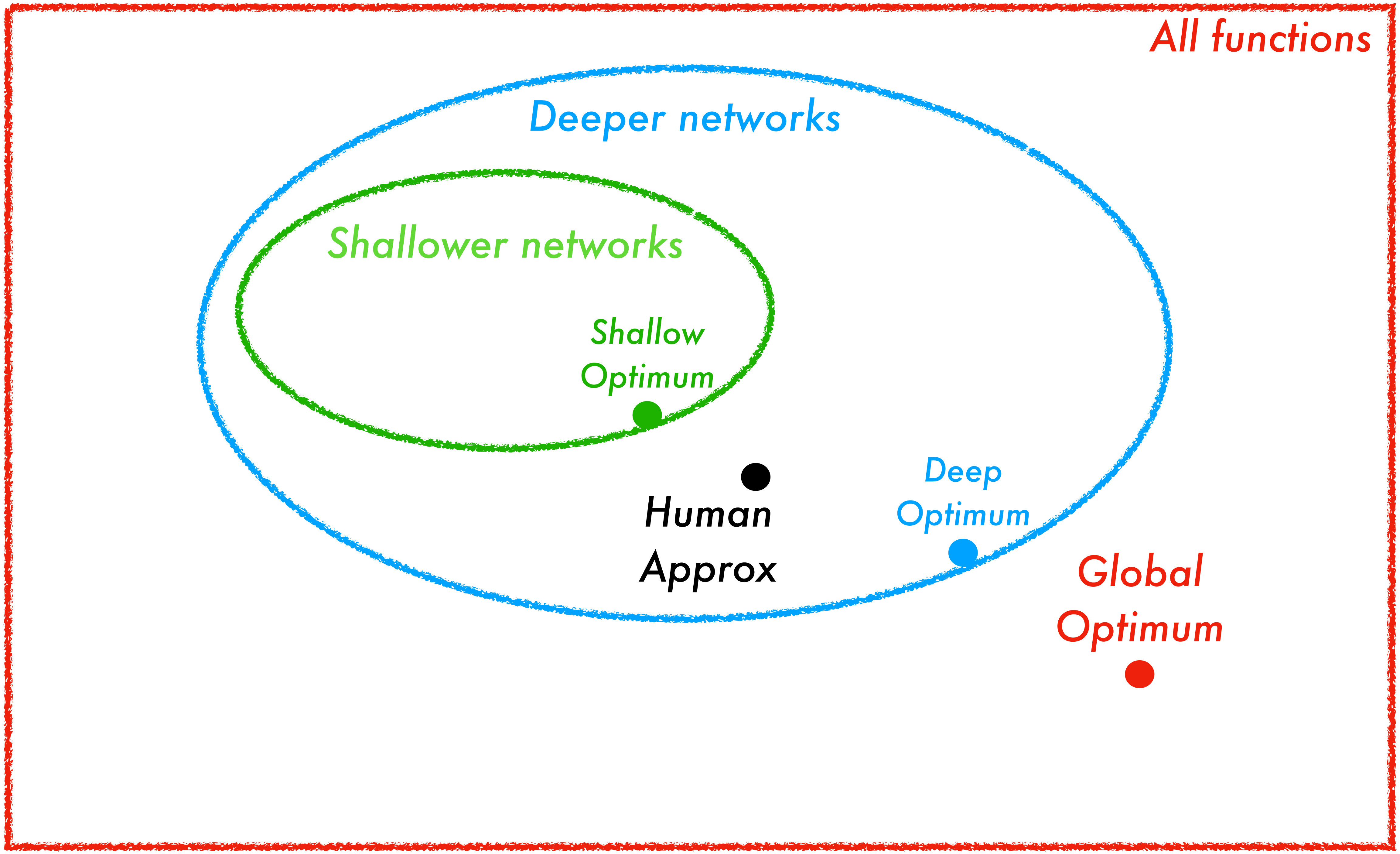
DOI: 10.1038/ncomms5308

Searching for exotic particles in high-energy physics with deep learning

P. Baldi[1], P. Sadowski[1] & D. Whiteson[2]

arXiv: 1402.4735

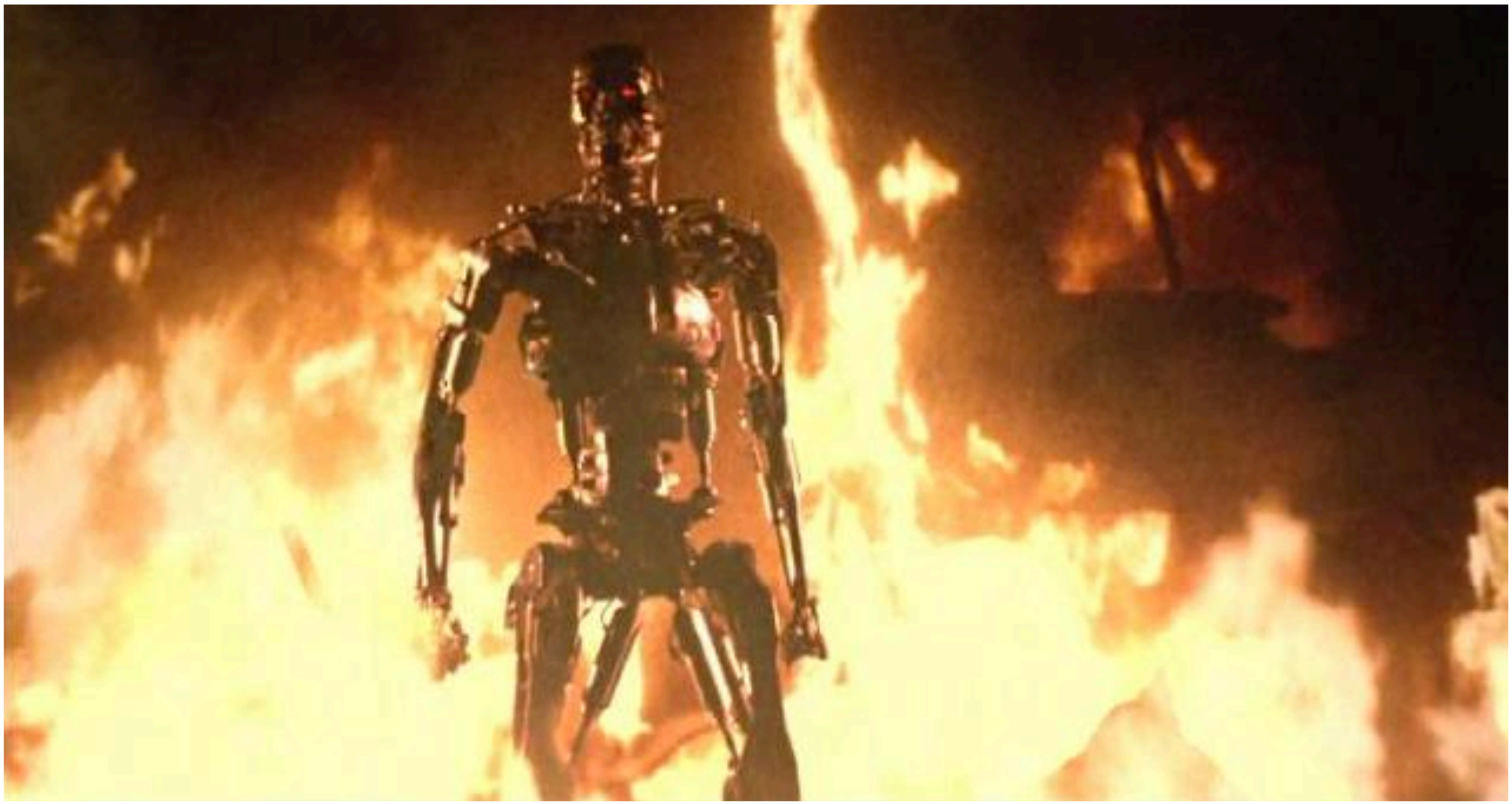
Expanding space



Results

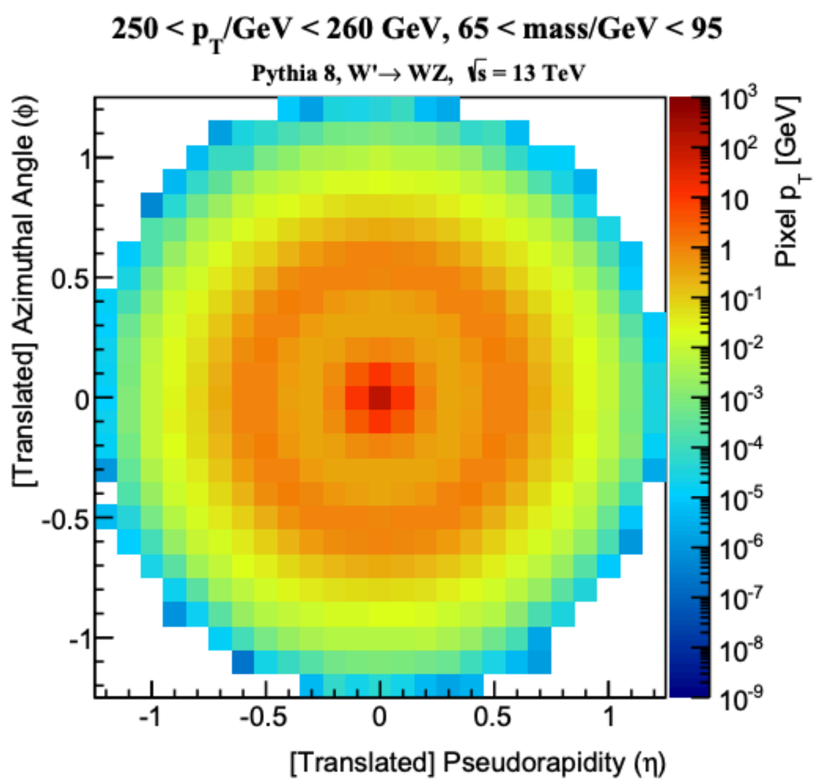
Deep networks succeed without human insight.
Outperform shallow networks and human ideas.

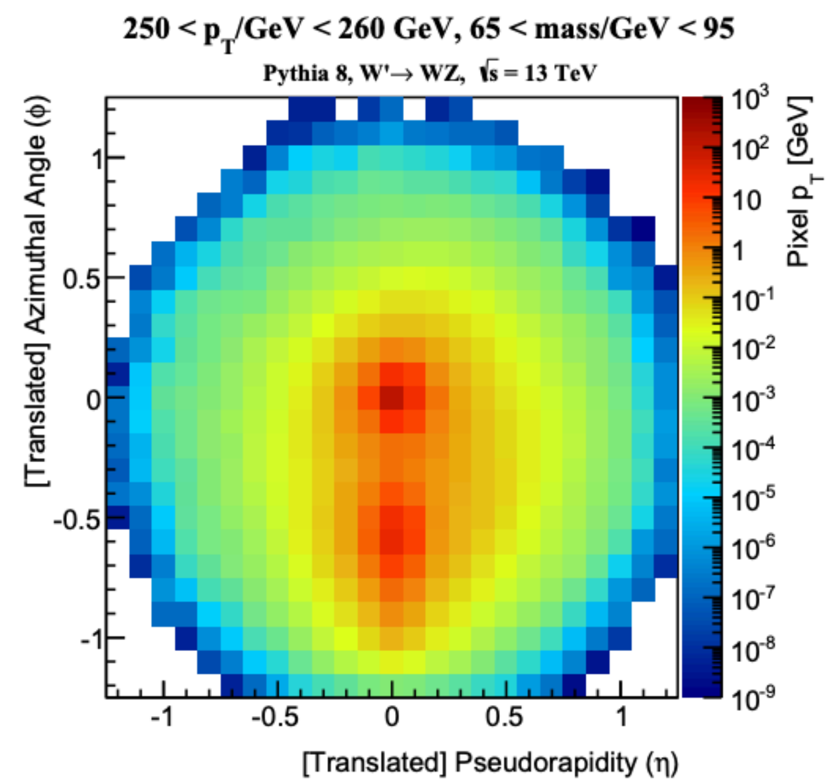
The AIs win



Low level data

Calorimeter pixels





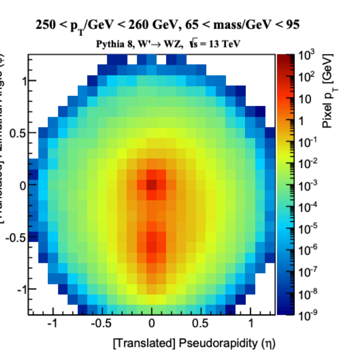
Networks
beat experts!

1511.05190 1607.08633

How?

What is it doing?

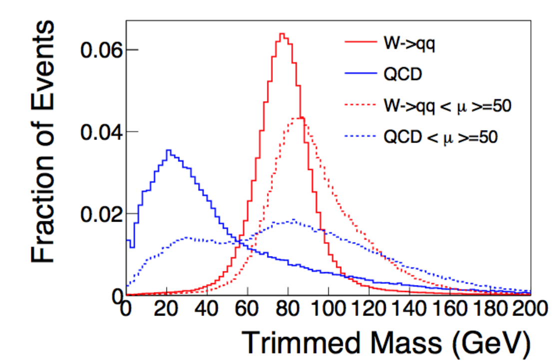
Our low-level (LL) data are often high-dim



LL

expert physics knowl.

HL



Can't interpret LL data

But HL doesn't always capture the information

Yet we prefer HL

<u>If HL data includes all necessary information...</u>

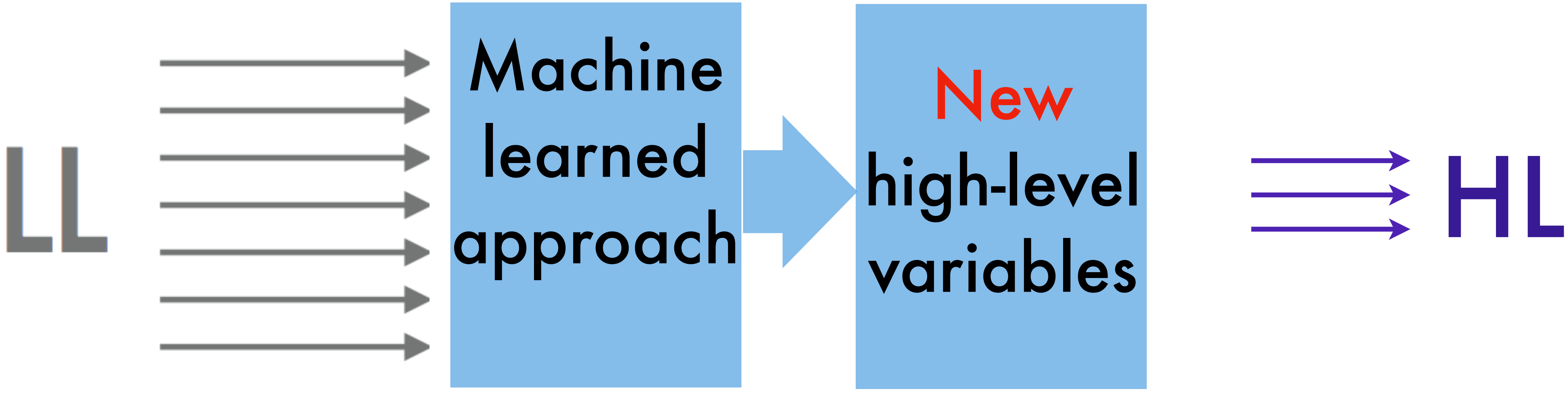
- It is easier to understand
- Its modeling can be verified
- Uncertainties can be sensibly defined
- It is more compact and efficient

- LL -> HL is physics, so we like it.

Our question

What can we learn from the machine?

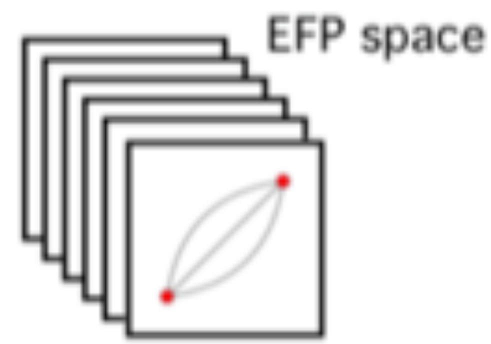
Learning from ML



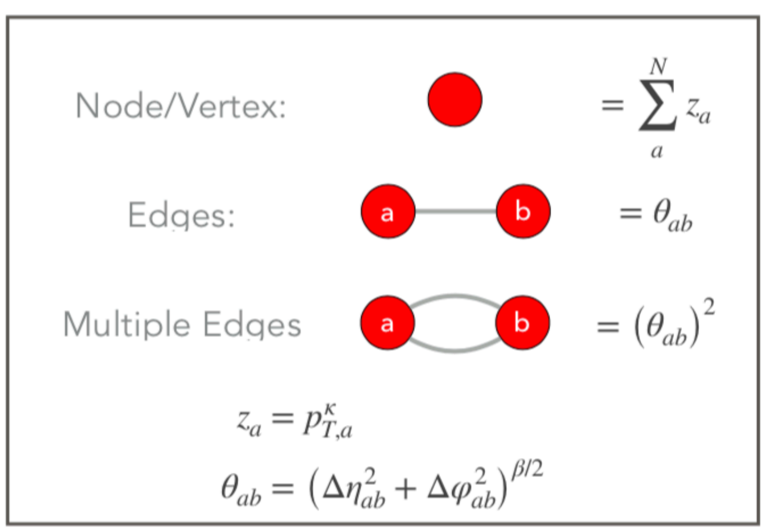
Use LL analysis as a probe, not a final product.

How?

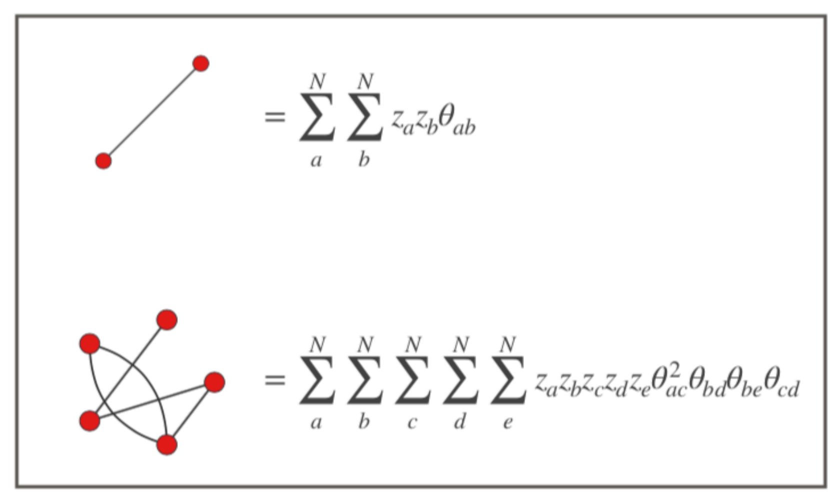
Define the language



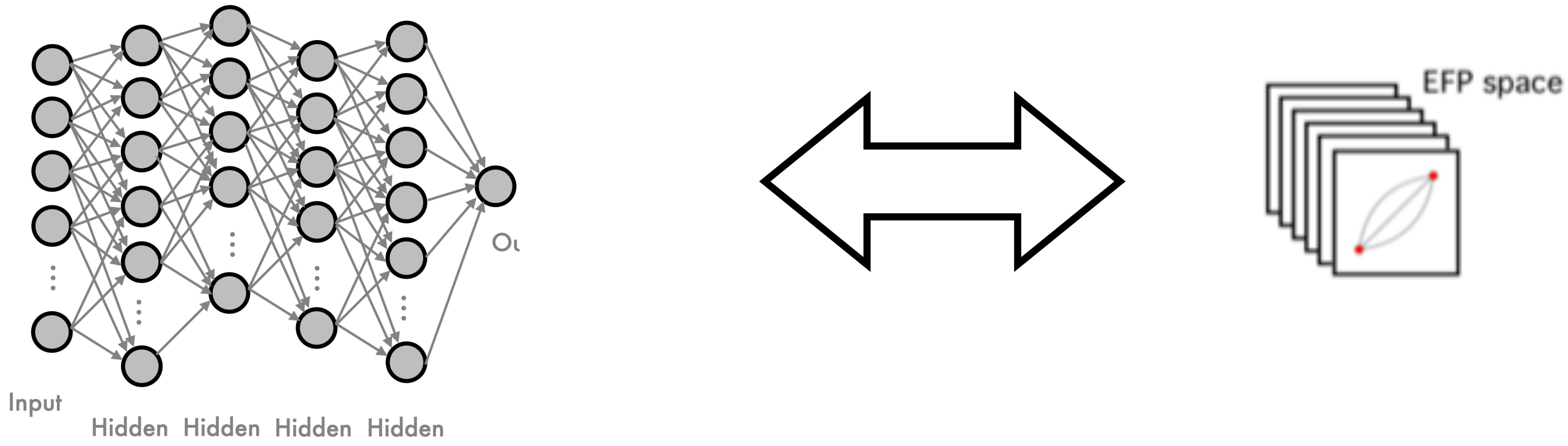
Graph components



Examples

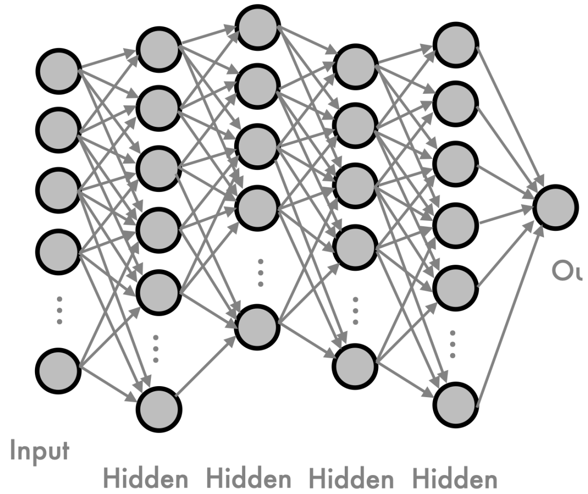


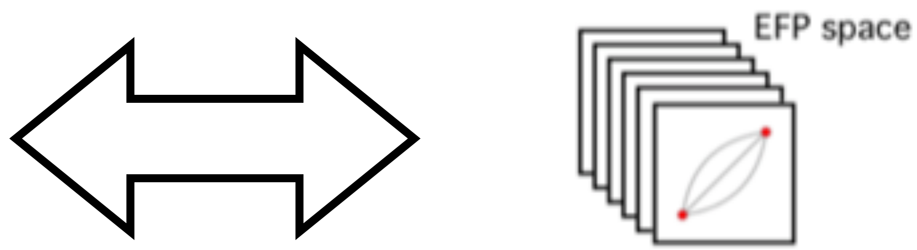
Mapping



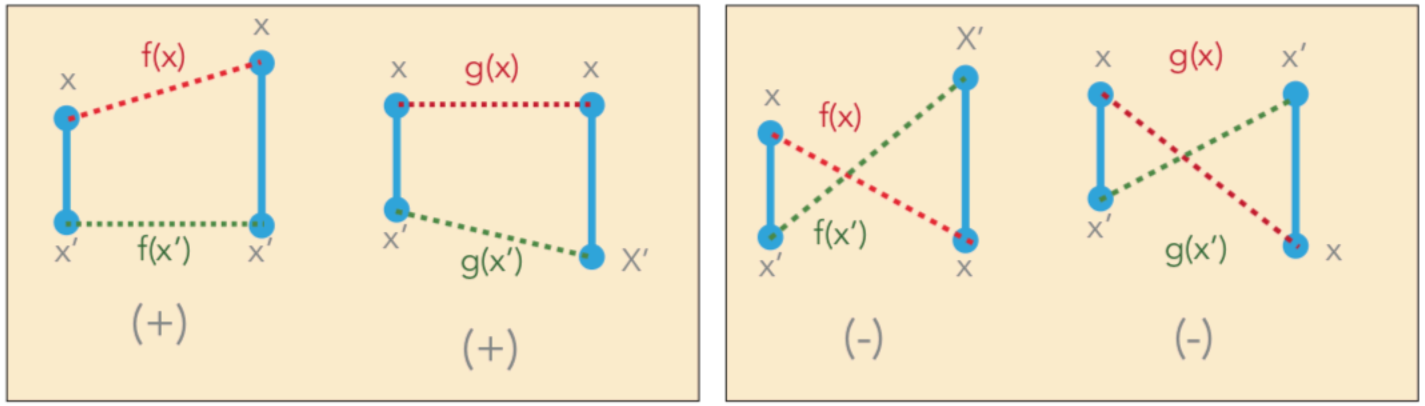
How to map from deep network
into our space of interpretable observables?

Mapping





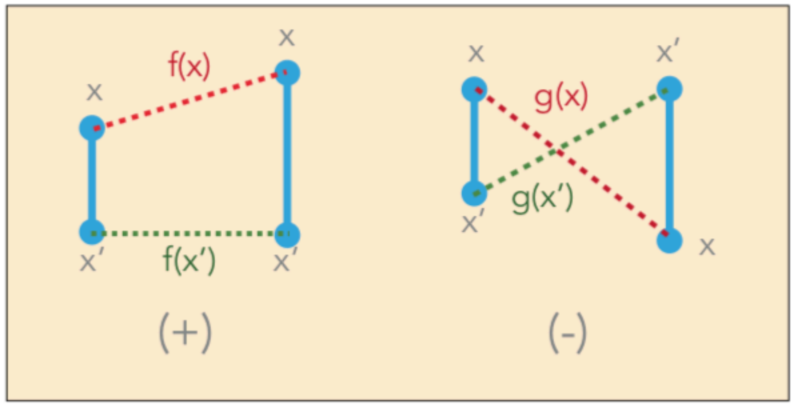
Similar Orderings



Function sameness

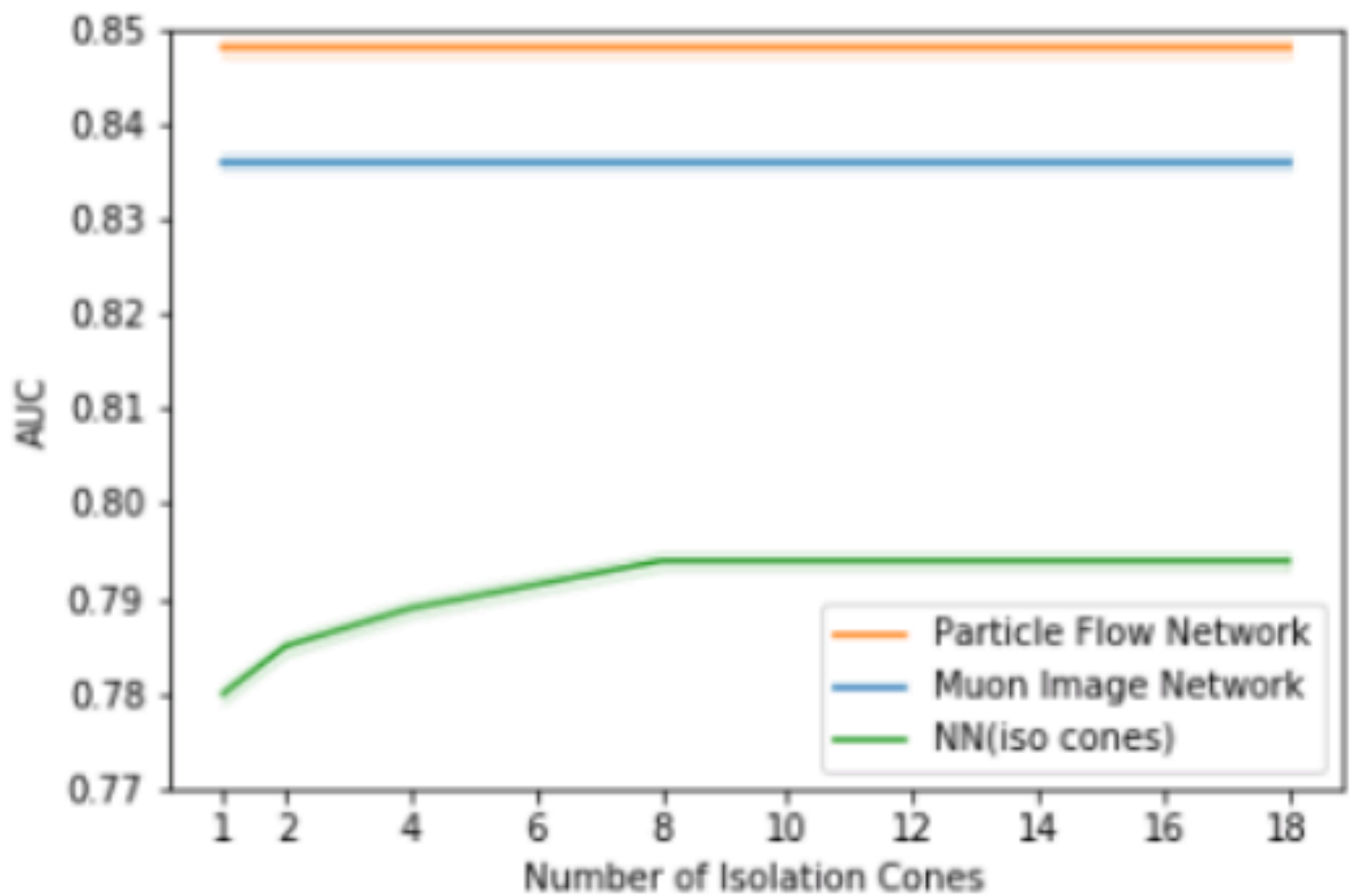
Complete equivalence not important

Dissimilar Orderings



Only care about the ordering of points

Results



Deep learning

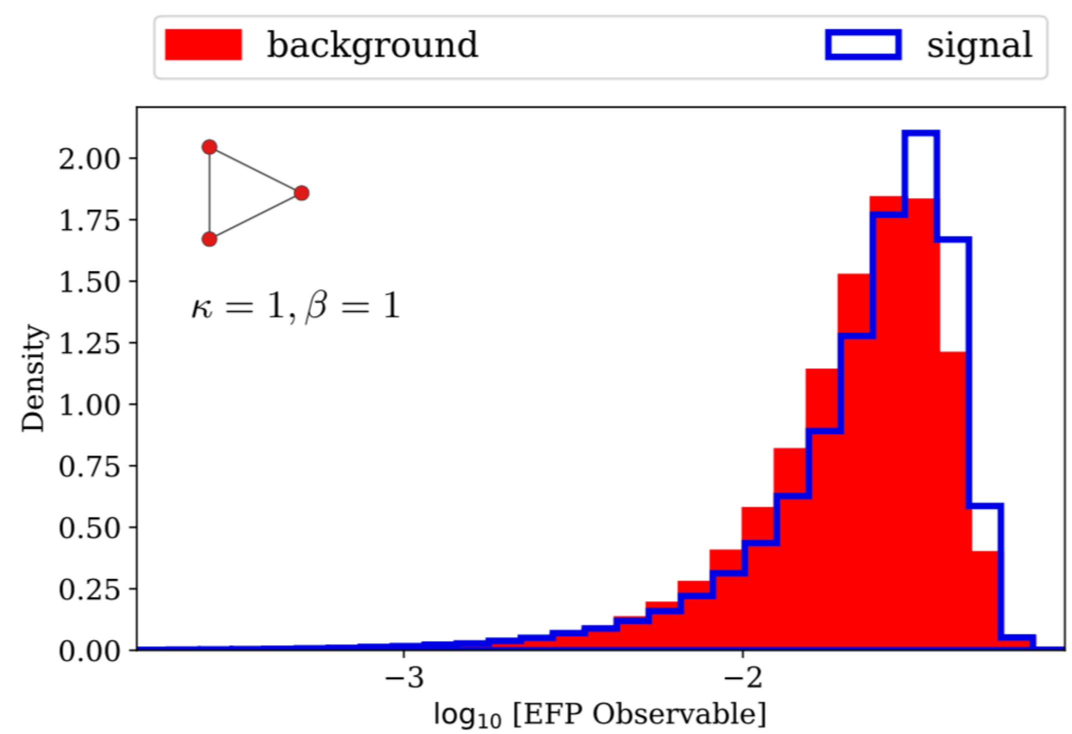
Human approx

More iso cones improves performance
Isolation cannot match calo-cell networks

Useful observable

This observable helps!

$$= \sum_{a,b,c=1}^{N} z_a z_b z_c \theta_{ab} \theta_{bc} \theta_{ca}$$



Procedure also works in real data
without any labels

Conclusions

<u>Deep Learning is a powerful new tool</u>

offers faster learning of nonlinear functions

<u>We have many appropriate tasks in HEP</u>

traditional heuristics should be re-examined

<u>No replacement for human intelligence</u>

garbage in will still give garbage out