Technical Report

On Probabilistic Rationalism

Wolfgang Rhode

December 17, 2020

technische universität dortmund

SFB 876 Verfügbarkeit von Information durch Analyse unter Ressourcenbeschränkung

Part of the work on this technical report has been supported by Deutsche Forschungsgemeinschaft (DFG) within the Collaborative Research Center SFB 876 "Providing Information by Resource-Constrained Analysis", project C3.

Speaker: Prof. Dr. Katharina Morik
Address: TU Dortmund University
Otto-Hahn-Str. 12
D-44227 Dortmund
Web: http://sfb876.tu-dortmund.de

Abstract

The question of how to arrive at scientifically secured knowledge has accompanied research from the very beginning. Depending on the scientific context and the historical epoch, answers have been given, essentially depending on the demands on the degree of truth and the scientific methodology. Recently, a new scientific methodology has emerged which is best characterized as "probabilistic rationalism". This methodology is the subject of this article: In cooperation between computer science and physics in the last decades and years, methods have been developed, allowing to analyze massive amounts of data collected in modern experiments under consideration of their probabilistic properties. Artificial intelligence or machine learning is the methodology of the day. The considerations presented here are based on the understanding of the statistical nature of the cognitive process, from which the probability character of scientific statements results. In Dortmund, we have been cooperatively engaged in investigating not isolated aspects of these analyses but the entire evolutionary process of knowledge expansion. In the present study, interdisciplinary aspects of epistemology from the perspectives of physics, computer science, and philosophy are combined to form an up-to-date and consistent model of knowledge acquisition. With this model, some known problems of existing epistemological approaches can be overcome. There are interesting parallels between the functioning of machine learning and biological-neuronal learning processes.

Keywords— Epistemology, Machine Leaning, Probabilistic Explanation, Astroparticle Physics, Models of Data

1 Basics, questions and motivation

One goal of epistemology is to answer how inferences from observations of nature, respectively, from measurements in experiments, can be drawn to explanatory theories. As all measurements are always a record of a finite number of discrete information (data points), there is no logical conclusion to the generalizing mathematical function of the describing theory. In the natural sciences, falsicationism, the basic concept of critical rationalism has therefore largely prevailed. In this context, from rationally conceived theories, functional connections between experimentally measurable quantities are predicted, which are being tested in experiments, attempting to falsify the postulated connection and with it the predicting theory (Figure (1)).

If falsification is successful, a new, better theory must be developed. If the experiment confirms the theory, the experimental tests should be continued under more stringent conditions. This epistemological model presented 1933 by Karl Popper in his book "The Logic of Scientific Discovery"[1] is based on clearly defined logical conditions. However, as stated in various discussions between Karl Popper and his critics, critical rationalism is neither intended nor suitable for describing physics research's real temporal course. Points of criticism against Popper arise just because of this logical structure when trying to clarify under which conditions a theory can or must be regarded as refuted, in the world of modern very complex experiments and theories. How are experimental results to classify if experiments yield inconsistent results or different experiments contradict each other? What practical requirements for experimental statements have to be met for

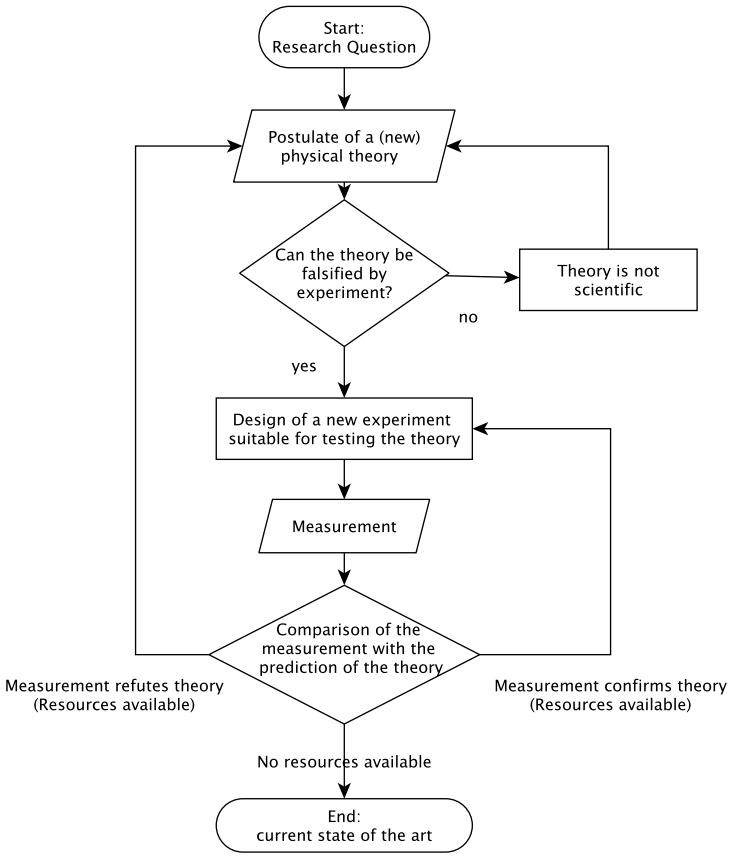


Figure 1: Connection between a physical theory and measurements in critical rationalism. The process will continue as long as the necessary scientific resources are available. Problems arise because no clear, logical answers can be given at the decision points in many real appearing cases.

a refutation of a theory? How exotic may the requirements for experiments be, so that a theory can still be considered as disprovable by experiments and thus be called scientific?

Let me illustrate the problem with a famous historical analogy. Epistemology has its origin in the first approaches of the pre-Socratics, from which Plato derived his image of the cave allegory [2][1]. In this parable, the recognizing humans are fixed on a bench in

[1] "AND now, I said, let me show in a figure how far our nature is enlightened or unenlightened: - Behold! human beings living in a underground den, which has a mouth open towards the light and reaching all along the den; here they have been from their childhood, and have their legs and necks chained so that they cannot move, and can only see before them, being prevented by the chains from turning round their heads. Above and behind them a fire is blazing at a distance, and between the fire and the prisoners there is a raised way; and you will see, if you look, a low wall built along the way, like the screen which marionette players have in front of them, over which they show the puppets. - I see. - And do you see, I said, men passing along the wall carrying all sorts of vessels, and statues and figures of animals made of wood and stone and various materials, which appear over the wall? Some of them are talking, others silent. - You have shown me a strange image, and they are strange prisoners. - Like ourselves, I replied; and they see only their own shadows, or the shadows of one another, which the fire throws on the opposite wall of the cave? - True, he said; how could they see anything but the

a cave behind a brick wall so that they only can see the cave wall opposite them. They cannot look back to see that the shapes dancing on the cave wall in front of them are the work of puppeteers who move their figures over the brick wall behind the researchers' backs and in front of a fire as a source of light, thus creating shadows (Fig. (2)). If one, like the group of philosophers added by Sanreadam in 1604 in the picture to the side of the wall, has cognition *a priori* of Plato's theory of what is going on in the background, then the *top down* explanation of the observed phenomena is simple.

Figure 2: Plato's allegory of the cave after Jan Saenredam, 1604 (c) The Trustees of the British Museum [3]

A fundamental epistemological problem is the question of whether there is a way to conclude backward, *bottom up*, from the effect (in the picture: "from the shadow") to its cause (contained in the physical/mathematical description of the mapping process). It was agreed that though in forward-direction[2] *top down* inferences from the cause to its effect can be made through logic or mathematics, *true* conclusions in the reverse direction back from the effect to the cause are not possible. When the Book of Nature was opened

shadows if they were never allowed to move their heads? - And of the objects which are being carried in like manner they would only see the shadows? - Yes, he said. - And if they were able to converse with one another, would they not suppose that they were naming what was actually before them? - Very true. - And suppose further that the prison had an echo which came from the other side, would they not be sure to fancy when one of the passers-by spoke that the voice which they heard came from the passing shadow? - No question, he replied. - To them, I said, the truth would be literally nothing but the shadows of the images. - That is certain. -"

[2]Which of the directions of inference between measurement and theory is perceived as "forward" and "backward" depends on the scientific community. In this article, the direction is defined to fit the mathematical nomenclature of the solution of the "inverse problem" that later turns out to be central. "Forward" thus passes inferences from theory to experiment.

by Galileo[3] both, experiments were described in the language of mathematics and measurements were accepted as starting points of inductive inferences. Then the physical explanation of the world began to progress rapidly and with great success. However, the central question remained unsolved, how this functional and, with time, increasingly adequate but heuristic description of the world could be justified epistemologically.

Newton demanded as a solution to this question that similar impacts must be caused by the same causes[4,5] Doing so, he was fortunate that this postulate applied to the gravitational effect he was considering in the sublunar world of the Earth's surface, and in the translunar world of the solar system, and that he was thus able to unify the phenomena of both physical worlds. The subsequently rapid and successful developing classical mechanics suggested that the world of nature could be mapped into a world of clearly solvable mathematical problems and theories. Also, the new path from the phenomena of electro- and magnetostatics to Maxwell's equations, though leading to a somewhat more complex structure of the descriptive theory, did not change anything about the epistemological classification of this form of explaining the world. Statistical statements and correlations forced in thermodynamics their way into the world explanation based on the ignorance interpretation of probability[6]. Later, in quantum mechanics, the more far-reaching demand for an identity of probabilistic structure and physical law turned against the realm of what could be said with certainty. At first, however, these developments seemed to be only disturbing artifacts within a deterministic explanation of the world.

The practical path from experiment to theory always seemed to include the necessity to abstract a few (preferably) true and clear statements from the experimental measurement. These propositions, in which an observed phenomenon (e.g., the spiral track of

[3]"Philosophy is written in this grand book -I mean the universe- which stands continually open to our gaze, but it cannot be understood unless one first learns to comprehend the language and interpret the characters in which it is written. It is written in the language of mathematics, and its characters are triangles, circles, and other geometrical figures, without which it is humanly impossible to understand a single word of it; without these, one is wandering about in a dark labyrinth."[4]

[4]Rules of Reasoning: (I) No more causes of natural things should be admitted than are both true and sufficient to explain their phenomena. (II) Therefore, the causes assigned to natural effects of the same kind must be, so far as possible, the same.[5]

[5]"Quest. 31. Have not the small Particles of Bodies certain Powers, Virtues, or Forces, by which they act at a distance, not only upon the Rays of Light for reflecting, refracting, and inflecting them, but also upon one another for producing a great Part of the Phenomena of Nature? For it's well known, that Bodies act one upon another by the Attractions of Gravity, Magnetism, and Electricity; and these Instances shew the Tenor and Course of Nature, and make it not improbable but that there may be more attractive Powers than these. For Nature is very consonant and conformable to her self. How these Attractions may be perform'd, I do not here consider. What I call Attraction may be perform'd by impulse, or by some other means unknown to me. I use that Word here to signify only in general any Force by which Bodies tend towards one another, whatsoever be the Cause. For we must learn from the Phenomena of Nature what bodies attract one another, and what are the Laws and Properties of the Attraction, before we enquire the Cause by which the Attraction is perform'd."[6]

[6]From a deterministic starting point, one concludes: "In principle, all processes in physics are exactly calculable. Because we cannot do this at the moment because of our personal mathematical incapacity or because it would be disproportionate to the desired result in terms of time, we are content with a merely statistical description of the phenomenon"

a particle in a cloud chamber at a known magnetic field) is interpreted as something ("the particle is an electron with an energy in the interval between E_1 and E_2") are those (basic) sentences, which are set up and checked by Popper according to the rules of logic, and which decide in critical rationalism via falsification or confirmation about the fate of preferably structurally simple theories with very few free parameters.

The current situation of knowledge acquisition in physics with data science methods differs in three aspects fundamentally from the situation described above.

1. We are no longer dealing with a few individual observations or measurements, but with the collection of such large and complex — albeit selectively collected — amounts of data that the compilation of a few individual statements, which could be classified by humans, is no longer conceivable. The analysis of a massive amount of data makes such different qualitative demands on its epistemological treatment that this whole can no longer be described effectively as a sum of its purely logical interpretable single statements.

2. What is electronically recorded or measured, in such large numbers, is simple. This are charges, times and positions – and nothing else. These measurements are only indirectly related to the dependencies (e.g., energy spectra, angular distributions, mass distributions, or regularities in their temporal fluctuations), which are to be investigated in the language of physical theories. Usually, most of the measured data is not only entirely irrelevant for the investigated phenomenon but even disturbing. In spite of the large amount of data recorded, the measurement space can be very sparsely occupied. Therefore, in addition to mapping the data into the theory, the selection of the subset of the data relevant to the physical problem to be analyzed has to be investigated first. In this view, phenomenon, and measurement are linked by probability distributions, maybe suspended at very few measuring points (like e. g. in radio astronomy).

3. A third difference to the classical view is that we are no longer the first to stare at Plato's cave wall. Whatever form of ratio in the past may have justified the path to today's physical theories, we already have a very sophisticated and successful model conception of the world fixed in mathematical theories. So the task is not to conclude from "something observed" to an explanatory relationship for the first time, but it consists in checking whether a model[7] is compatible with the measurement in the light of all known contexts. Today, this can be done with statistical (Monte Carlo) simulation calculations, in which in one virtual reality the physical knowledge about the investigated process is mapped with high precision and under consideration of all probabilistic parts of the problem description into the charges,

[7]This model can, for example, contain the standard models of particle physics and cosmology and other theories confirmed so far. It is used for the explanation of the cosmos, which is supplemented by new effects of, for example, dark matter, dark energy, sterile neutrinos or other not yet excluded phenomena of particle physics.

times, and positions to be registered in the measurement. Today's data analysis aims to statistically classify the difference between the measured values and the values expected from the theory and translate this difference into physically relevant statements. This task is handled through statistics and no longer by Aristotelian logic.

2 Excursion: Physical Problems

In general, the epistemological considerations sketched here do not only apply to the treatment of questions from a sub-area of physics, but they claim general validity. However, this article is aimed at readers with different scientific backgrounds, so that it seems appropriate to present a paradigmatic case of physical practice to which the examples given later may refer. For completely subjective reasons and without any restriction of universality, I choose my working field of astroparticle physics as a starting point. Astroparticle physics is an approximately three-decade-old field of research [8] that has emerged from the extension of astronomy by experimental methods and theoretical concepts from particle physics. Its understanding is impossible without the current theoretical standard models of particle physics and cosmology and their possible extensions. Thus, it is based on the classical field theories (for gravity) and quantum field theories (for particle interactions), which have not yet been unified. The essential components of astroparticle physics are shown in Figure (3).

In astrophysical sources (such as stars, supernova remnants, the nuclei of active galaxies, and many other objects), temperatures are so high that matter must be considered as plasma (atomic nuclei and electrons are separated and move independently). The charged particles cause irregular magnetic fields and are vice versa accelerated by these magnetic fields towards very high energies. The electrons and nuclei from the plasma can interact with other particles or emit radiation so that the sources can finally be observed from the detection of the light of electromagnetic radiation between the radio and gamma range. Also the neutrinos produced in hadronic interactions in the sources can be detected. A general glow in the sky of astroparticle physics is produced by the charged nuclei, which mostly lose their directional information when flying through the different magnetic fields on their way and are therefore of limited use for astronomical purposes. After repeated interactions, the follow-up products of astrophysical accelerations can be detected with different techniques using satellites, earthbound telescopes of different designs, and underground detectors. These individual and globally distributed detectors also use techniques developed in particle physics. To complete the knowledge about the astrophysical sources, the information from all detectors must be combined. At the latest, after this combination, the different components of the cosmic rays must be separated and understood.[8]

[8]For further discussion see [9].

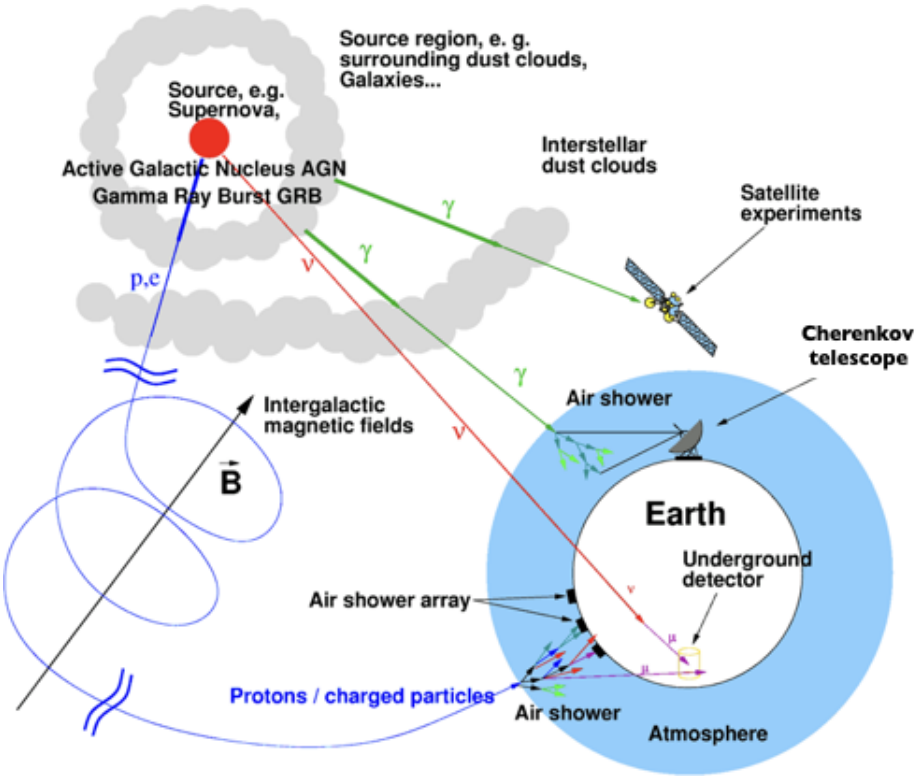


Figure 3: Sketch of astroparticle physics. Indicated are the astronomical sources, the messenger particles and the different detection scenarios (changed following [7]).

Depending on which of the three components contributing to the measurements (source properties, radiation propagation between the source and Earth, or interactions of elementary particles) is considered unknown, insights can be gained in astroparticle astrophysics, cosmology, or particle physics. Because of the complexity of the required physical descriptions, the (anyway neither necessary nor purposeful) assumption of ***completely*** controllable initial conditions cannot be made in this field at all. What is detected is the result of a concatenation of processes described by different theories and phenomenological approximations. In the end, these results may be precision measurements, e.g., to close measurement gaps between experiments, or to test predicted signatures, or to discover unexpected new phenomena that can only be explained by an extension of physical concepts. In the three decades of its existence, astroparticle physics has contributed to the revision of outdated theoretical ideas through many observations, despite or because of its complexity. And perhaps it is therefore well suited for the discussion of epistemological boundary conditions of current physics using the science of big data.

In the historical development of physics one finds repeatedly dependencies between new physical insights and the new mathematical methods applied for them. The method of machine learning was developed in parallel to astroparticle physics. And it took some time until the early insight of computer science into the necessarily probabilistic nature of analytical results [11] was also accepted in physics.

3 From the physical theory to the physical observable

The statistical methods used in modern physics, together with machine learning methods developed in computer science, provide a scientific toolkit that allows the basic epistemological question to be answered consistently. The core of these methods consists of determining those probability densities needed to establish connections between the world of registered values and the world of formulas in theoretical physics. To emphasize the idea pursued in this paper, it is necessary first to discuss methodological facts and steps. It will be shown in the following that what at first sounds like "methodical tricks" actually contains the key to new aspects of epistemology. The principle of this form of knowledge acquisition consists *top down* in forward-direction for deriving phenomena from causes in creating a complete virtual reality in the simulation, including all known fluctuations and oscillations based on the existing physical knowledge. On the one hand, the predictions concerning the measurement results calculated in this way can be directly compared with experimental measurements.[9] On the other hand, this virtual reality, in which the (virtual) truth and the (virtual) measurement results are always known at the same time, can be used to check the validity, efficiency, and robustness of the applied inference methods of machine learning to be used in backward-direction, *bottom up*, inferring from the data to elements of the physical theory. The paradigm is outlined in Figure (4).

To understand the situation, we first consider the *top down* calculation of an experimental effect resulting from postulating a function $f(x)$ to be investigated in the frame of a physical model. The specific probability distributions of the data to be detected and registered by a suitable detector are then to be calculated.

A generic detector: We collect the experimental data electronically and hence free of subjective influences such as reading errors. This fact also means that everything we can measure must ultimately always result from electromagnetic interaction, i.e. from the existence of charges or their change. The interaction of charges with the detector is read out electronically and (if necessary) stored. Such a measurement in the form of electronic detection of charge in a specific part of the detector will always take place at a particular point of time, e.g., at the end of a defined time interval in which the charge was previously deposited in that sensitive volume. This way, it can be compared, for example, whether a resulting voltage to be determined lies within an interval of reference voltages. The interval width of the reference voltages correlates with the measurement accuracy.

[9] This comparison could show that measurement and theory do not fit each other which would then be a starting point for a technical search for the reasons. Are there numerical problems? Is the experimental technique adequately described? However, no deeper *bottom up* inferences from experiment to theory can be justified from this direct comparison. See below (4.3) at solutions of the inverse problems.

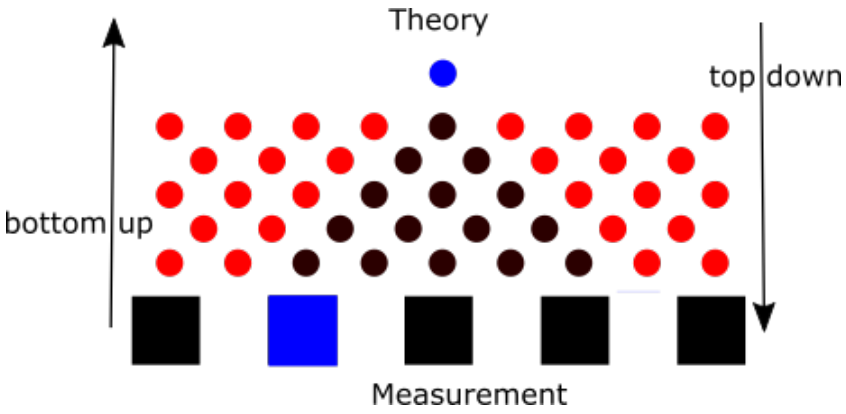


Figure 4: Sketch of the relationship between theory and measurement in a simple model: If an initial situation is given in the theory (initial position of the blue sphere), its fall through a generalized "Galton's nail board" can be calculated *top down*. At each of the black (according to Galton) or red (in generalization) nails, the sphere is deflected to the right or left until it might land in one of the square detectors or get lost in between. In statistics, the Galton's board is used as an illustration for the motivation of the Poisson distribution or, in the limit case, the Gaussian distribution, because of its simple structure. The probability distribution of signatures for a given theory value can thus be calculated with presupposed theoretical knowledge. However, only the experimental signature (sphere in blue square) is given after a measurement. Reverse to this, the epistemological problem requires determining the probability distribution for an initial value formulated in the nomenclature of theory based on the measured signature *bottom up*. This calculation is possible if methodical boundary conditions are met - contrary to the cave parable's trivial view. The epistemological meaning of the calculation and these boundary conditions is explained in this paper.

Using electrical oscillators, high-precision clocks can be built. If necessary, these clocks can be used so that the sketched readout process takes place in very short time intervals.[10] Thus the chronological sequence of charge measurements in the intervals represents the *bin-wise*, *digitized* measurement of time-dependencies.

A detector usually consists of a large number of sub-detectors (pixels). In each of these sub-detectors, the described processes take place. We know the position, form, and size of these electrically sensitive components. Thus a spatial image of the observed phenomenon can be created for each time interval. Because of the sub-detectors' discrete nature, this image is also spatially divided into intervals, *pixeled*. Neither the temporal nor the spatial intervals, in which the detector is sensitive, must fit together without gaps. A loss of signals is always possible. The theoretically possible signals will be registered only with a certain probability.

[10] In practice, this can take place today with high frequencies of more than 10^9 Hz, i.e., more often than one billion times per second. For the argument, the size of the frequency, resp. the time interval of the measurement is irrelevant.

Therefore, the result of a measurement is always a spatially and temporally discrete sequence of (integer) numbers. Its relative accuracy depends on the design of the experiment and the data acquisition electronics but is limited in any case. The exact amount of charge recorded electronically will always depend on thermal, problem-related numerical, or quantum mechanical fluctuations, which will create differences in the digitized results. Finally, there will always be several (maybe unwanted) physical possible causes to change the detector's charge state. In any case and beyond other causes, the desired signals are also influenced by electronic noise.

Everything on which our measurement of the world is based, everything that we can have as an unquestionable result of a single measurement, are thus multi-dimensional, spatially and temporally discrete frequency or probability distributions of the measured charges, positions, and times. To make statements based on such measurements possible, the measurements must be repeated until the shapes of the probability distributions are known with sufficient accuracy. In an idealized form, we can write a measured probability distribution as a continuous function $g(y)$, taking into account that the actual measurement result consists of the discrete elements of a vector $\vec{g}$, which contains in each element i the normalized sum of the entries between the minimum and maximum y_i, thus containing the normalized frequency distribution.

The connection between theory and measurement: The connection between a prediction of theoretical physics[11] $f(x)$ and the experimental result $g(y)$ is given by the formula (1), here for simplification only noted one-dimensional and continuously. According to our initial analogy, this formula states nothing else than an expression of Plato's cave equation:[12]

$$g(y) = \int_a^b A(y,x) \cdot f(x)\,\mathrm{d}x + b(y) \tag{1}$$

In this Equation (1), x is the variable of a physical theory whose connection with the measurement must be examined. For example, x could be an elementary particle's energy whose existence and properties must be investigated in a detector. Measured are the integer numbers y (charges in time intervals at specific positions). After a sufficiently large number of the measurement repetitions, a frequency distribution of $g(y)$ containing the (many) single measurements of y can be specified. The occurrence of a x is a discrete act, either because it is discrete (decay of an atomic nucleus leading to the emission of the elementary particle under investigation) or because the intensity of x is averaged over the time measurement interval.

[11] This can be a single value, a prediction in an interval or an unlimited function.

[12] In terms of Plato's cave, the $g(x)$ corresponds to the shadows on the wall, which result from the unknown reality $f(x)$ transformed by an also unknown imaging procedure $A(y,x)$ and which are disturbed by a unknown number of unwanted artifacts $b(y)$.

An occurrence of x thus leads to a measurement of a state y. Between both worlds of theory and measurement results, there can be many levels of theoretical description ("Galton's nail boards in all imaginable forms and irregularities"). For example, we consider a telescope to detect high-energy neutrinos from various sources: If the first particle here is a muon-neutrino, it could, in a weak interaction, generate a muon which, as it propagates, reaches the detector and, via various forms of electromagnetic interaction, deposits energy there, part of which is emitted as Cherenkov light. These photons could pass through the detector if transparently designed. They may be scattered or absorbed until they cause the photoelectric effect in the detection instrument, which in turn may trigger electrons that - after amplification - can be detected by the measuring electronics.

All of the above processes must be understood in theory with high precision. This fact presupposes that this theoretical understanding is based on well-confirmed background knowledge which has been established by other experiments. Since stochastic elements of thermodynamics, quantum physics, or quantum field theories sooner or later appear in the ladder of all descriptive theories (as in the example), even a fixed x will always produce different y with a frequency distribution $g(y)$ characteristic for the measurement process, if the measurement is repeated several times.

As already discussed in consideration of the generic detector, we can assume that each measurement is repeated very often, so that always ensembles of measurement results $g(y)$ have to be considered. The single measurements central to the classical epistemological discussions in the philosophy of science, or the significance of single measurements' analysis, would have to be treated as special cases. The components of the formula (1) can thus be explained as follows:

- $\mathbf{g}(\mathbf{y})$ is a frequency or probability distribution **as derived from very many repetitions of the measurement**. This distribution represents the result of the measurement. Since the data is recorded electronically, it can only be based on distributions of positions, charges, and times or quantities calculated from them. This frequency distribution of the measurement results is additively composed of two components of different meaning:

- $\mathbf{b}(\mathbf{y})$ is the contribution of **unintentionally measured components**. This fact refers to signals that are recorded only for technical reasons because it is *a priori* unknown whether, e.g., a change in the detector's charge state is caused by the phenomenon under investigation, or by another reason not related to the investigated question. The cause may be some form of electronic noise, or it may consist of various signals that can be physically analyzed and are important for answering other questions but which do not play any role concerning the specific question currently under investigation. This contribution, potentially consisting of different components, is called *background*. This background can be derived from theory or measured directly with the detector (e.g., with the signal switched off, if possible). That $b(y)$ must be a frequency, or probability distribution is given by the technical reasons already discussed.

- $\mathbf{f(x)}$ is the **frequency distribution of the variable x to be inferred in the experiment** and causing the measurement results. The variable x is written as part of the world of physical theories in physical units and must therefore be translated into the world of electronically recorded positions, charges, and times or the quantities calculated from them, in which $g(y)$ is given. This translation depicts the understanding of the measurement process. Fixed values for x are conceivable but barely practically realizable. Experimentally nearly always, at best, a very narrow probability density around the desired value can be realized.

- The **translation of a single value x into a contribution to the measured frequency distribution** $g(x)$ is performed by the so-called design function $\mathbf{A(y, x)}$. This function depicts all physical properties of the experiment, including the subdetectors. It summarizes the sequence of all interactions and their stochastically generated results (in the example, all that lies between the energy of the primary neutrino x and the frequency distribution $g(y)$ registered by the electronics). Only in very rare and simple cases $A(y, x)$ can be given analytically as a solution of nested integrals. How this function can be numerically determined is explained below in the section on Monte Carlo calculations.

- By the integral $\int \mathbf{A(y, x)} \cdot \mathbf{f(x)}\,\mathbf{dx}$ all contributions containing the **relevant information** for the examined question are summed up to the measured distribution. To calculate the contribution, the causal theoretical function $f(x)$ is folded with the design function $A(y, x)$, which translates the values from $f(x)$ into the measured units of $g(y)$. The integration must be done via the probability (resp. frequency) $f(x)$ of all possibly contributing x (e.g., via the energy spectrum of the neutrinos to be detected) to obtain the sum of all contributions to the signature $g(y)$ to be measured.

If the spectrum $f(x)$ is known, the expected distribution of the measured values $g(y)$ can be derived if the design function $A(x, y)$ and the background $b(y)$ are also known. Interpreting formula (1) as a modern notation of Plato's cave allegory, the mapping of theory into experiment is described from an omniscient perspective. From this perspective, the experimental result can be assigned to the theory since all functions involved are known. If a light source and the illuminated object were known, the shadow and, if wished, its digital image could be calculated. If the blue sphere position above the Galton's board were known, the frequency with which it stays in this or that pot detector could be calculated. This procedure describes the inference from cause to effect.

Monte Carlo simulations: As evident from the example of the neutrino detection, the previous discussion shows that the central problem is shifted to finding the design function $A(x, y)$. For practical purposes, it cannot be expected that this function can be given by analytically solving the large number of nested integrals that describe the different stages of the interactions. In the search for a numerical representation of the design function, however, one encounters a method known as *Monte Carlo* to solve integrals by stochastic evaluation.

To illustrate this method, it can be imagined that an unknown area and a known area enveloping the first one are evenly covered with randomly distributed points. The number of points on both surfaces can be counted and their ratio can be calculated to determine the unknown surface.[13,14] The method is called *Monte Carlo* because of the random scattering of the points. By a high number of randomly distributed points, arbitrarily exact results can also be obtained on sub surfaces. This is relevant for the mentioned nesting of processes. What at first seems to be only a halfway smart numerical trick to the solution of nested integrals, on closer inspection, opens up a multitude of epistemologically relevant possibilities. Therefore, the world of Monte Carlo simulations is now inspected in more detail.

For this purpose, again, as an example, the propagation of the elementary particle muon is discussed: Muons are "heavy electrons" whose interaction probability is so low that with the appropriate starting energy, they can penetrate kilometers into materials such as stone or water. Various processes (the most important of which are ionization, pair formation, bremsstrahlung, nuclear interaction) may occur. Each of these processes takes place with a specific probability (depending on the energy and material). It has specific properties regarding the amount of energy loss suffered by the muon and the muon track's deflection during an interaction. The corresponding probabilities have been calculated in quantum field theory. On their way through matter, muons, like spheres on Galton's board, suffer a multitude of corresponding interactions [12, 13].

Because of the originally stochastic character of the (quantum field) interactions, a muon's path cannot be determined deterministically. However, some of its properties could be calculated by suitable averaging and integration over all the probability distributions involved. These are, e.g., the mean position, the mean velocity, and the mean deviation of the muon from a straight path after crossing a certain thickness of the material, and other quantities. This form of calculation corresponds to classical approaches and the attempt to produce testable predictions [10]. Apart from the effort involved in this procedure, the method has the disadvantage that it can only provide the centers and perhaps the moments of the resulting probability distributions for the investigated properties ($g(y)$).

[13] There are more elegant numerical ways to perform the method in practice.

[14] A circular surface can be calculated (as a simple example) so that a square with an inscribed circle is evenly covered with equally distributed random points. To determine the circular area, one counts the points in the square and the circle and knows the ratio of the number of points to the area.

However, as discussed above, we are interested in the complete probability distributions of specific signatures. This desired result is obtained by randomly generating individual muon trajectories using the Monte Carlo method: In our example, it is possible to determine first, based on the total quantum field theory, interaction probabilities and derive where the next interaction will occur. Then, which of the partial cross-sections mentioned above (pair production, bremsstrahlung, ...) will be employed. Then, how much energy will be released in the material, and then at what angle the muon continues its path. According to this pattern, the possible muon trajectory can be calculated step by step until the muon stops or decays.

Generating that way large numbers of possible muon trajectories, the initial condition (muons start at the position x in direction $\vec{a}$ with the energy E) can be correlated with, e.g., the distribution of the charges, times, or positions y_i at which (some of the) muons hit the detector, or simultaneously the number of hit subdetectors y_j. This is always and independently of all the individual properties of a specific path possible. Geometrical peculiarities (different materials at different positions, geometrical boundary conditions, etc.) can easily be integrated into the process so that, finally, very realistic simulation results can be generated. Always a distribution of the resulting y is assigned to a primary x. The design function $A(x, y)$ is thus obtained by determining the distribution of the final y assigned to a primary x.

From integrals to matrices A further, initially only technical step proves to be relevant for understanding the cognitive process, because it prepares the inversion of the direction of cognition (*top down –> bottom up*). So far, we have written the cave equation as convolutional **integral**, thus taking into account the continuous character of theoretically calculated probability distributions. However, due to the necessarily discrete nature of the measurement results, Equation (1) can also be written as Equation (2) in the language of linear algebra:

$$\vec{g} = A\vec{x} + \vec{b} \tag{2}$$

The vectors $\vec{g}$, $\vec{x}$ and $\vec{b}$ contain in each element the number of entries counted in the respective specific interval as binned frequency distributions. The size of the intervals is arbitrary and can be chosen depending on the problem. In $\vec{b}$ the contributions of the background are counted, in $\vec{g}$ all contributions to be measured and in $\vec{x}$ the initially mostly continuous function $f(x)$ is integrated bin-wise from the lower to the upper interval boundary. The translation of numbers with physical units into numbers in detector units is done here using the design matrix **A**, in which (as described) all Monte Carlo results are assigned and counted. In this picture, the properties of the matrix **A** can be understood:

- An ideal detector would register every particle crossing its geometrical volume and measure observables for all its properties with absolute accuracy. The determination of the particle energy in such a detector, e.g., could work via the number of cells hit by the particle and, per definition, perfectly correlate to the energy. The matrix **A** is then a diagonal matrix where the probabilities contained on the main

diagonal (besides normalization factors) would be equal to one. All other entries are equal to zero.

- A detector may be crossed by some particles that hit its mechanical structure but not its sensitive volume. If the existing measurements would be absolutely accurate, the matrix **A** would also be represented by a diagonal matrix. Because of the lower probability of reconstruction, the probabilities in the main diagonal are then less than or equal to one. All other entries are zero here as well.

- If in addition, the measuring accuracy is not perfect, i.e., if several possible heat numbers are assigned to **one** energy (maybe because the track lengths of the particles with the same energy in the detector and so the number of hit cells vary *also* by geometrical reasons), the elements next to the main diagonal begin to populate with entries unequal to zero. This range becomes the wider, the probabilities outside the main diagonal become the larger, the worse the resolution (the higher the probability for misallocations) in the detector is.

The Matrix **A** is thus determined as the result of the Monte Carlo simulation by counting the combinations of the physical quantity x (in the example: the energy) and the detector quantity y (in the example: the number of hits in a subdetector) in each element of the matrix. A high number of simulated entries in one element means that this number was determined with a low statistical error. Few entries mean that this combination occurs very rarely and is therefore only known with a relatively sizeable statistical error. After normalization to the number of generated events, the matrix contains probabilities.

In summary, it can be stated so far that single measurements produce singular entries to the concerned probability distributions. A large number of measurements determines the shape of these distributions. With numerically highly accurate Monte Carlo simulations, based on all known physical knowledge, the expected measurement result can be calculated for a known function $f(x)$. The whole complicated procedure can be summarized by a simple linear algebra relation.

In trivial cases, where measurement and prediction do not match, a falsification criterion could already be applied here. However, the real problems in research are - as will be discussed in the following - so complex that the possibilities of knowledge acquisition are not adequately exhausted with simple binary decisions.

4 From the measured variable to the measuring point

Remembering the goal to conclude from measurements to theories, one may read Equation (1) in the integral form in backward-direction: $g(y)$ is then measured, $\mathbf{A}$ and $b(y)$ are then calculated or measured, and $f(x)$ is unknown and to be determined. With this status of knowledge, the unknown physical function $f(x)$ can be inferred from the measured numbers. In this reading, Equation (1) is a Fredholm's integral equation of the second kind. Because the described process in forward-direction is a convolution integral, the solution of this inverse problem is also called deconvolution.

The equation in its algebraic form (2) symbolically states the steps necessary for a solution of the problem. Therefore these mathematical steps shall be discussed separately before they are put into the final context of the multi-dimensional analysis of large amounts of data with methods of machine learning:

1. **Monte Carlo calculation of the design matrix A**: As explained above, to determine the matrix $\mathbf{A}$ first the simulation problem in forward-direction has to be solved. For this purpose, it is necessary to write the problem so that the simulation only includes correlations that can be assumed to be known for the present analysis.[15] To determine $\mathbf{A}$, a function $\hat{f}(x)$ (or a resulting vector $\vec{x}(\hat{f}(x))$) is assumed, which serves only to determine the correlation between $\vec{f}$ and $\vec{g}$ using the Monte Carlo method. The function $\hat{f}(x)$ chosen for this purpose is in principle arbitrary and only for computational reasons (resource-saving) it is advantageous to choose a form for $\hat{f}(x)$ corresponding approximately to the expected result $f(x)$. We can consider the simulation problem solved with the previous explanations.

   From an epistemological point of view, it is essential to notice that the reversal of the direction of the inference also changes the direction in which the matrix $\mathbf{A}$ is read. For inferences from cause to effect, $\mathbf{A}$ determines which probability distribution of experimental signatures $\vec{g}$ is assigned to a dedicated theory value x. Here, for the inferences from effect to cause, $\mathbf{A}$ assigns to a dedicated measured signature g the probability distribution of possible theoretical causes $\vec{x}$. This applies as well to the probability that a measured signature is triggered by a signal event, as by a background event. Furthermore, this is also valid for the probability with which to a measured signature (hit count) the size of a property from the world of physical theory (the energy) can be assigned.

2. **Background subtraction**: To subtract the background $b(y)$, it must either have been simulated with Monte Carlo methods or alternatively been measured with the

[15] Of course, the input for every simulation input is always subject to uncertainties from different sources. How these uncertainties affect the results is always examined and considered in the special analyses, but should not distract our thinking. As a result of such systematic investigations, one finds that to the *statistical uncertainty* from the number of measurements comes a further *systematic uncertainty* from ignorance of the exact conditions and correlations. The extend of this ignorance, however, is assumed to be estimated on the basis of previous experiments.

detector under special conditions. Then it can be (symbolically) subtracted from $g(y)$ in Equation (2):

$$\vec{g} - \vec{b} = \mathbf{A}\vec{x} \tag{3}$$

The distinction between signal and background in real measurement data will be discussed below.

3. **Solution of the Inverse Problem**: The goal to calculate the vector $\vec{x}$ from the measurement of $\vec{g}$ would be fulfilled by multiplying Equation (3) with the inverted matrix $\mathbf{A}$ from the left:

$$\mathbf{A^{-1}}\left(\vec{g} - \vec{b}\right) = \vec{x} \tag{4}$$

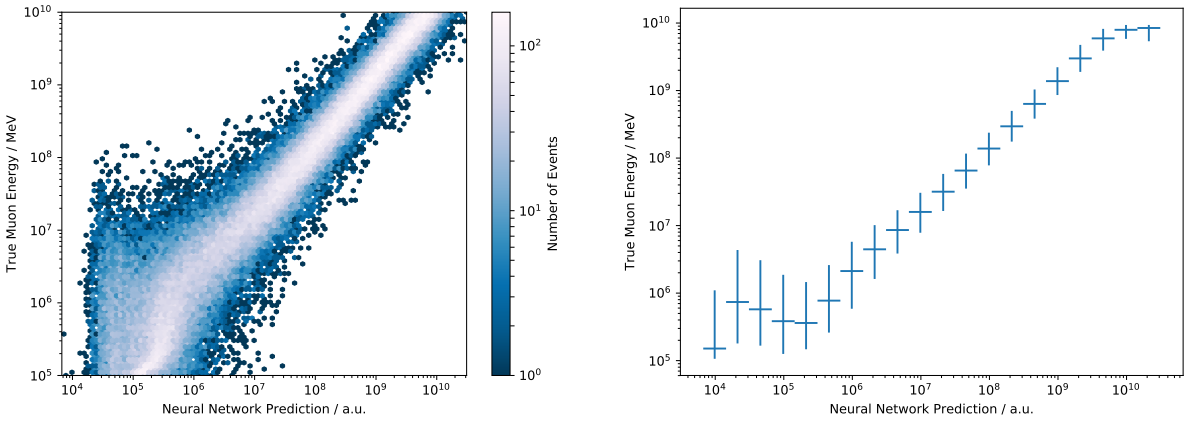
Apart from the mathematical difficulties that can occur with matrix inversions, one central problem remains to be discussed: When calculating the matrix $\mathbf{A}$ we have seen that its elements are based on the simulation of large numbers of entries. Near the main diagonal of $\mathbf{A}$, there are numbers close to probabilities of 1 with small statistical uncertainties, derived from many entries. Far away from the main diagonal, elements with only a few entries generate probabilities close to 0 based on a small entry number and correspondingly large statistical errors. An example of correlations for the construction of a realistic matrix $\mathbf{A}$ is given in Figure (5), where the estimation of the energy of muons as it might appear in a generic neutrino telescope is presented.

"Measured" in the simulated detector of Figure (5) are times and charges at many positions for 100000 muons with different energies. Then, using a trained neural net[16], this multidimensional information is projected to a one-dimensional quantity in arbitrary units correlated to the energy. Shown is the typical correlation between the output of the neural net on the (x-axis) and the logarithm of the true neutrino energy in the simulation (y-axis). A given muon energy can lead to different neural net outputs, and vice versa, a certain neural net output can have been activated by different muon energies. This is a necessary consequence of the thermodynamic and quantum physical processes involved. The correlation with its statistical consequences via matrix A is considered in the procedure discussed here.

On the left of Figure (5), the frequency of the occurring combinations is color-coded. Rare single entries are visible as single black points. Below it is shown, that they cause the ill-posedness of the problem. In classical analytical approaches, one would have tried to identify a unique calibration function fitted to the mean-values and width of the distribution of the correlations, depicted

[16]Neural nets as machine learning algorithms will be discussed below. For the moment we understand the "trained neural net" as somehow invented function to calculate numbers more or less correlated to energy.

right in Figure (5) for the same data. Such a calibration function, however, leads to false ideas about the unambiguousness of the assingment between the derived magnitude (neural net output) and the searched-for physical magnitude.



(a) Color-coded correlation of combinations between a calculated neural net output and a physical quantity (muon energy).

(b) Binned mean values of the left frequency distribution. The error bars indicate the width of the distribution.

Figure 5: Example: Estimation of the energy of muons as it might appear in a generic neutrino telescope. Shown is the inference from a neural net output (x axis) to the true energy (y axis). The difference between a calibration function (right) and a frequency distribution (left) for the same problem is obvious. The Monte Carlo Simulations were executed with the lepton propagator PROPOSAL [12, 13, 14].

By inversion of the matrix A, large numbers become small numbers and vice versa. Nevertheless, large uncertainties remain large and small ones remain small. Thus single, random entries in $\mathbf{A}$, irrelevant in the forward-direction, might become large, noisy terms (large number with considerable uncertainty) after inversion, disturbing a clean determination of a result. Therefore the class of these inverse problems belongs to the so-called ill-posed problems, where small numerical differences in the argument (single random entry in $\mathbf{A}$) can produce large differences in the result. As in the calculation of $1/x$ close to zero, this is independent of the mathematical formulation of the problem and the way of the problem solution. This unpleasant situation can be cured by suppressing the disturbing terms in $\mathbf{A}$, called regularization. The regularization is based on additional assumptions about the properties of the solution to eliminate the fluctuations. Thus, a similar but not ill-posed problem is created. However, that way, via the regularization, necessarily "prejudices" about properties of the solution enter the solution. Therefore, for every practical problem, the impact of the regularization on the solution of the problem has to be investigated. The regularization must not change the physical interpretation of the measurement. This can be quite critical if one is looking for new physical effects that cause a "sudden jump" of the physical function to be measured (e.g.,

the change of an energy spectrum by the generation of a previously undetected particle, causing a "peak" in the spectrum) because such jumping of functions is just the expected for poorly regularized solutions of ill-posed problems.

We consider the measurement of the function $f(x) = N \cdot x \cdot e^{-ax}$ in 20 measuring points (bins), as an example. Instead of using simulations to determine the detector properties, we assume a simple illustrative Matrix **A**: our measurement is performed with a detector that does not lose any signals but assigns them with a probability of ϵ (in a friendly manner symmetrically) to correlated matrix elements. If the probability for mismatches ϵ would be $= 0$, the matrix would be diagonal and the detector ideal. The matrix is shown in Equation (5).

$$A = \begin{pmatrix} 1-\epsilon & \epsilon & 0 & 0 & \dots & 0 \\ \epsilon & 1-2\epsilon & \epsilon & 0 & \dots & 0 \\ 0 & \epsilon & 1-2\epsilon & \epsilon & \dots & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & \dots & 0 & \epsilon & 1-2\epsilon & \epsilon \\ 0 & \dots & 0 & 0 & \epsilon & 1-\epsilon \end{pmatrix} \tag{5}$$

In Figure (6) the function $f(x)$ to be measured is shown as a solid line. If **A** would be inverted and the Equation (4) applied, the strongly oscillating blue measuring points might be obtained because this is an ill-posed problem. If the relevant uncertain terms are moderately suppressed by regularization, the green measuring points are derived fitting very well to the initially assumed function. A too strong regularization, however, forces the (now violet) measuring points more and more to a straight horizontal line.

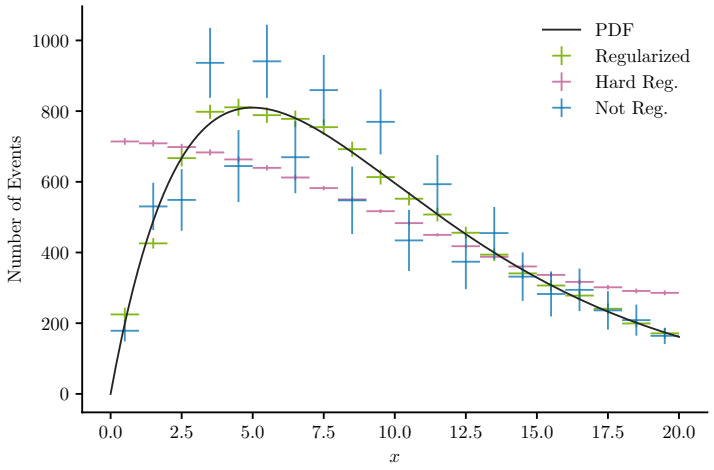


Figure 6: Example for the solution of the inverse problem without (blue measuring points), with appropriate (green measuring points) and with too hard (violet measuring points) regularization.

Epistemologically it is to note that inferences from the effect on the cause lead to inverse problems. Inverse problems are by their nature ill-posed and can produce fluctuating solutions. The fluctuations can be suppressed by regularization, i.e., a suitable variation of the mathematical function to be solved. Inevitably, however, it is then necessary to weigh physically possible fluctuations of the measurement result against the bias strength of regularization as a smoothing condition of the solution.

Multidimensionality: Before discussing the solution of the problem with means of computer science, the limitation of the observed quantity to one dimension shall be removed. With each measuring act in an actual detector, many measured variables are electronically read out. Further quantities are calculated from these original numbers. Analogous to the above discussion, the central data analysis question is how probability densities for the physical properties of the investigated events can be calculated from (highly) multi-dimensional measured signatures. Not only the measured data but also the properties to be reconstructed in the world of physics will usually be multi-dimensional (although in fewer dimensions).

Signal-Background Separation: As the first step in data analysis, the signal events of interest must be filtered out of the entire data set (subtraction of the background b, see above). Since the unwanted but technically unavoidable background can often be many orders of magnitude larger than the searched-for signal, this indicates the search for a tiny needle in a gigantic haystack. In the exemplary astroparticle physics data analyses, the background may exceed the signal by up to 10 orders of magnitude. In the example, a wrong decision may only be made less than once in a trillion events. Also for less ambitious problems, fast, highly precise, and reproducible decision procedures are required, in which also the error rate is calculable. Such tasks can no longer be solved with the necessary precision by human decisions. Instead, they require the use of high-precision robust rating algorithms, developed in the frame of machine learning.

These **machine learning** rating algorithms are optimized according to to-be-investigated quality criteria. First, the signal background separating task is defined in the world of simulations. In this world, it is known which multi-dimensional signatures occur with which frequency and which of these signatures are assigned to the signal respectively background class. To this purpose, many examples for the sought signal events and all possible variants of underground events are generated with Monte Carlo methods.

The goal of machine learning is to identify mathematical operations to be applied to the many dimensions of generated and, in principle, measurable variables (number of hits in the detector, their spatial distribution and temporal sequence, size of the deposited charge, etc.), assigning finally a value on a one-dimensional finite scale to each event. The variables, the mathematical operation, and their internal structure are optimized to maximize the distance between signal and background events on that scale. For this optimization, simulated signal and background events are successively evaluated by the algorithm. The larger the number of these examples is, the better the algorithm can be

optimized. Anthropomorphically, the process of entering an example is called "showing," and the process of optimization is called "learning".

Thus, the algorithms are optimized for separation tasks. Since the necessary decisions are statistical processes, the accuracy with which a decision is made must be specified. Furthermore, it must be tested whether the algorithm behaves sufficiently tolerant to unavoidable imperfections of the simulation compared to the experimentally determined data. If one is finally sure about the procedure, one knows how many simulated signal and background events are expected in the data set after applying the separation algorithm, including the uncertainties of these numbers. Applying the so-determined algorithm to the world of measured data, one can transfer the selection rates determined in the simulation to the world of measurements.

In the research area of machine learning, many algorithms have been developed in which representations are structured with entirely different approaches to map multi-dimensional measurement results to one-dimensional order parameters.

Depending on how many free parameters have to be calculated to solve the task, the algorithm can, on the one hand, better or worse recognize structures and is, on the other hand, more or less dependent on the precision and the number of Monte Carlo events. Only a few free parameters are used in the so-called *discriminant analyses*, in which for the separation task ***flat***, n-1-dimensional hyperplanes are optimally placed in the n-dimensional space of the measured signal and background distributions. Distributions that lie curved and nestled into each other are obviously difficult to separate with this method.

Conversely, in so-called (deep) Neural Networks, many parameters are available to describe even fine structures. In extreme cases, optimization of such algorithms can finally lead to a *too good* separation in the simulation world, in that sense that the positions *of all* generated events are parameterized respectively "learned". If applied to real data or even to another simulation data set, such a network would fail because it does not find signatures to be *identical* to those learned with the simulation.

Which machine learning approach is chosen and how it is validated that it delivers adequate and expected results, has, therefore, to be analyzed separately for each separation task.

As exemplary examples of such algorithms, first simple decision trees are considered. A decision tree is a hierarchical set of conditions that defines that a data set (event) is classified in one or another form if the measured parameters lie in specific logically nested intervals. In a single decision tree with a small number of conditions, very simple classifications can be executed. The more complicated the signal-background separation is, the more conditions would have to be nested and inserted into the tree. These conditions,

however, are determined based on the simulated Monte Carlo events. If a sufficiently large number of conditions is selected, individual properties of the Monte Carlo events are mapped, and thus the classification of all simulated events could be stored. In this case, the algorithm would not be able to learn generalizing the topology. Such a generalization would be indicated by a robust parameterization of the placing of the searched for signal-background-separation plane. Without this generalization, the separation will fail in all practical applications.

Alternatively, a so-called *Random Forest* can be applied. In a Random Forest, an ensemble of decision trees with moderate branching possibilities is defined instead of one a widely ramified, deep tree. Not all trees of the Forest are provided with the same information. The trees are further only given randomly selected information (some simulated observables plus the label showing the true class affiliation). Also, on which branching level which information is used in the decision process is determined randomly. After the numerical optimization of a Random Forest with simulated events (called "training"), all trees of the forest "vote", by calculating their scale value for an event answering the question, whether an event belongs to the signal or background. This results in the above required one-dimensional scale, on the one side of which the trees ideally vote for "underground" (scale value close to 0) and on the other side, the trees vote ideally for "signal" (scale value close to 1). A cut on this scale can then be used to determine the ratio of signal to background events, the purity of the finally selected signal sample, the efficiency of the selection (How many signal events survive the cut?), the robustness (Is the Forest as desired insensitive to e.g. Monte Carlo fluctuations?) or in the agreement between Monte Carlo and measured data. Finally, the decision process of the specially trained and tested Random Forest is executed on the measured data. As desired, a sample is obtained in which the number of signal and background events and their statistical uncertainties are known.

As required, one finds that way a mapping from the high-dimensional space of the (real or simulated) measurement data to a one-dimensional order parameter. The multi-dimensional character of the measured data is preserved and is supplemented by the order parameter calculated by the Random Forest. Considering the measured variables as carriers of a multi-dimensional probability distribution, the signal-background separation can be understood as a transformation of the multi-dimensional probability distributions for the signal and background samples enabling a separation of the both topologies by a simple linear cut. The strength of a selected signal can now be given as the probability that the signal is inconsistent with a random fluctuation of the background. This signal data set can be selected and used for further analysis.

Solution of the inverse problem: After the signal-underground separation, a data set may be assumed, containing only a known small number of background events that may be assumed as negligible for our purposes. In the example, one would now be sure to have a data set consisting of signatures of the sought-for neutrinos as signal events. Now, these signal events are to be examined for other properties. For example, this could concern the reconstruction of the energy spectrum of the neutrinos.

The paradigm of signal-background separation consisted in the requirement that signal and background are perfectly known in all properties except the relative number of events in the classes and that they, therefore, can be described just as correctly in the simulation. If, however, a quantity such as the energy is to be measured, then (at least) the energy distribution of the singular events cannot be wholly known. Therefore the simulation cannot describe the distribution to be measured later perfectly. For the deconvolution, a weaker paradigm is postulated, according to which only the experimental effect of the searched phenomenon has to be known. Only the simulation of the detector in all of its properties (e.g., the response to the energy) must agree in nature and Monte Carlo.

Regardless of this, the unfolding analysis is based on the same data situation in measurement and simulation as the signal-underground separation: We consider a multi-dimensional spatio-temporal pattern of positions, times, and charges and variables derived. The Monte Carlo data set for each event contains additionally the true information about the quantity (energy) to be measured, which is unknown in the world of experimentally recorded data. Again a mapping from the multi-dimensional space to a one-dimensional parameter is sought. And again, this is a task that can be formulated mathematically as an optimization task: the simulated examples are used or "shown" one after the other to a machine learning algorithm to adjust the inner parameters of the learning algorithm so that the sought-after quantity (in our example: the energy of neutrinos) can be determined with the best possible accuracy. In place of the bipolar finite scale for the separation, now a continuous and unlimited magnitude (e.g., correlated to the energy) is calculated (compare Figure (5) and the example given there). The primary considerations for machine learning, however, will not change. Suppose a relation between the many technical quantities (charge, time, position, and derived from that) and the quantity (energy) to be measured physically is established. In that case, the inverse problem (Equation (4)) is solved under a controlled addition of assumptions to suppress unwanted properties of the solution (regularization). The analysis optimized in the Monte Carlo world is then equally applied to the measured data. If necessary, the signal to be measured is corrected for signatures that could not leave a trace in the detector. Also, a normalization of the event numbers considered so far to the physical theory quantities, including their unit is performed (by a so-called acceptance correction).

Thus, the deconvolution leads to a projection of the multi-dimensional measurement data into the one- or few-dimensional space of the quantities predicted in theoretical physics. Since at no point the statistical character of the simulated or measured data was limited or cut off, the expectation values and the corresponding confidence intervals can be specified for each measuring point in physical units. The result of the analysis are measuring points with their statistical errors. If a definite prediction for a measured value was calculated in theoretical physics before the measurement, it is now possible to specify the probability with which this prediction and the measurement match. If the theory contains parameters that cannot be predicted, they can be adjusted according to the physical measurement points (fitted). These parameters can then be introduced into further calculations and be tested there again.

As **epistemological lesson**, in a generalization of Popper's approach of falsification, we see that - even if special results are obvious - only probabilistic evaluations of measurements and theories are appropriate for the further discussion of the results and the consequences.

Remark on the structure of theoretical physics: Since the basic ideas of the "Logic of Scientific Discovery" were formulated, thinking about theoretical physics has changed. Originally, the change from Ptolemy to Copernicus to Kepler to Newton and finally to Einstein could be seen as a development that suggested the goal of elegantly developing theories using new approaches and mathematical methods and adding only a few measurable new parameters. Thus, this seemed to be a paradigm for the further development of theoretical physics. Today, whole classes of theories are proposed and calculated simultaneously. These classes of theories may contain a very large number of free parameters with possibly very large uncertainties, so that a complete refutation of such a class of theories is hardly possible. Modern epicycles, if you will, are built into the structure of the theory from the outset in such a way that they need not stand out as additional constructions. Nevertheless, there are of course parameter ranges within the classes of these proposed theories which can be meaningfully investigated by experiment.

This does not prevent understanding theories as products of rational considerations, and the goal of these considerations may still be to develop a mathematical description of the world that is as complete and uniform as possible. However, since a class of theories can rarely be fully tested, it would be a very hard demand[17] to link the scientific validity of a theory to its complete refutability. An experimental test of parts of the parameter range, however, might be possible.

Which parameter range is experimentally testable in which combination with others can often only be determined by means of Monte Carlo methods because of the large number of free parameters in the theories and the possible complexity of its dependencies: In these approaches, numerical values for the free parameters are randomly pulled, based on which possibly experimentally verifiable values for e.g., the resulting mass of a potentially existing dark matter particle or the flux of particles (neutrinos) from a class of active galaxies can be identified. The Monte Carlo method is thus suitable for testing the range of potentially occurring measurement results for two reasons. First, the relationship between the parameters and the potential measurement results is not always obvious. On the other hand, parameter regions excluded by other experiments can easily be considered.

[17]With this demand one would remove, e.g., many extensions of the Standard Model of particle physics from the field of scientific research, because they contain parameters which are not measurable in this world

The progress of knowledge: In the view presented here, the knowledge is expanded iteratively and evolutionarily: cyclically, the definition of the problem, the simulation of measurement results based on existing knowledge, the search for deviations from it in the experiment, the interpretation of the measured probabilities in terms of theoretical physics, and finally the revision of the problem follow each other (Figure (7)).

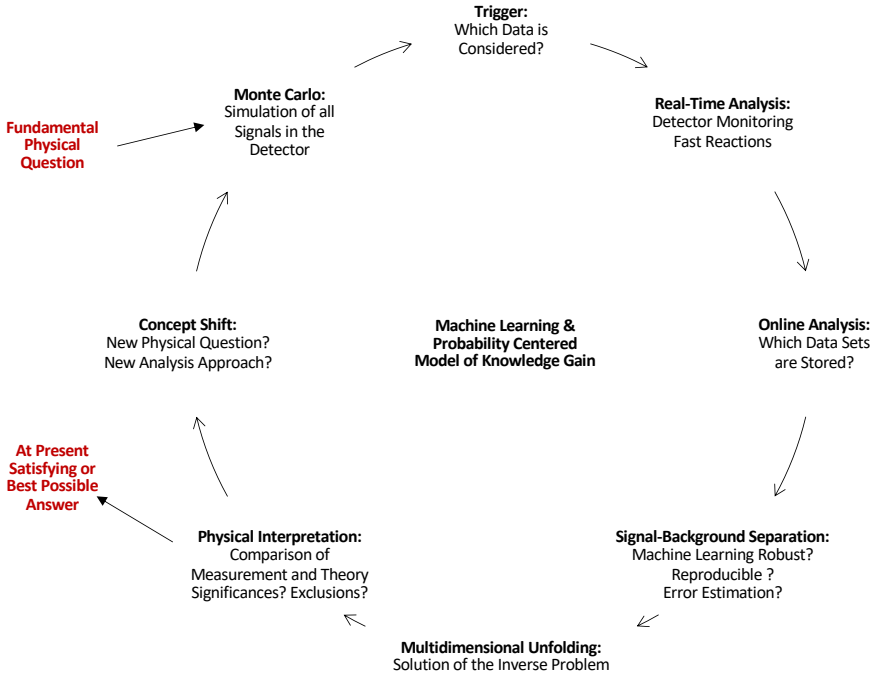


Figure 7: Cycle of methodical steps for knowledge expansion. Indicated are the most important aspects under which the multidimensional probability distribution based on the measurements is investigated and transformed for quality assurance and knowledge gain.

The acquisition of knowledge is that way no longer based on complete inferences from the effect to the cause, but rather on playing through all theoretically conceivable possibilities and constantly comparing and optimizing these predictions using new measurement data. The concrete and solid empirical basis on which is built for this purpose are the measured positions, times, charges, and the quantities calculated directly from them. To interpret these numbers, machine learning models are optimized or calculated in the discussed form. They provide probability distributions for the physical parameters measured in

experiments. In turn, the physical parameters limit the abstract space of possibilities found in the sea of theoretical physics models. Often it turns out that the available data do not give a preference for one or the other theoretical explanation. Occasionally, however, it is possible to exclude certain theoretical branches with a high degree of probability.

Automation: The preceding considerations show that the first solution of a problem is not possible without human intervention. Once solved, selection and reconstruction problems can be solved automatically over and over again. Based on this, the researchers involved in knowledge acquisition must now perform dedicated tasks. Theoretical physicists will develop new concepts for interpreting observed phenomena, constraining the parameter range. Experimental physicists will invent new techniques to enable qualitatively new measurements and further optimize existing techniques. And the computer scientists will develop new algorithms and methods that will allow the ever-increasing amount of data to be analyzed more accurately with fewer and fewer resources.

Remark regarding learning in biology: It should be noted here that the process described has formal analogies to learning in biology: also, biological learning is usually based on the evaluation of a large number of spatially and temporally discrete and electromagnetically induced neural information. From this information, possible relevant signals must be extracted by structure recognition and separated from irrelevant information. If necessary, the relevant information has to be further classified. Due to the consideration of many identified examples (e.g., observed animal species or plants or other things), necessarily blurred models are formed within certain limits (concept formation), which can be improved by identifying differences to the original model ideas. Properties of these classes are abstracted and arranged quantitatively. And also, in this kind of learning, it is less about absolute certainty than about sufficiently high probabilities, as represented in machine learning, e.g., by the voting behavior of trees of random forests.

This point of view is complementary to the approach initially motivating the development of the so-called Neural Networks and today still forming a narrative for the introduction of the method: Following this, attempts were made to reproduce the structure and function of neurons mathematically – but in fact, special methods for mapping multi-dimensional information into one- or few-dimensional order structures were developed. Today, we see that the understanding of the paradigm depends less on the mathematical form of the method than on its general functionality. The evaluation of this functionality as a problem-dependent quality criterion, however, is crucial.

Apart from this formal aspect, the understanding of the way in which machine learning works as paradigm to biological learning, offers interesting possibilities for clarifying long-examined questions such as the functioning of human thinking.

5 Conclusions

A probability-based approach to knowledge acquisition was introduced as an epistemological paradigm. Prerequisite for the application of this procedure is the calculation of the probability distributions of the experimental observables with Monte Carlo methods (simulation of a virtual reality) for all relevant physical effects. Thus it is shown that an experimentally testable combination of parameters from the world of physical theories induces a multidimensional digitized probability distribution of measured parameters. This probability cloud must first be transformed within the virtual reality so that methods to separate the relevant information from background or noise can be determined. Then, a method has to be established to invert the reading direction of the probability distributions so that for virtual experimental data, the probability distributions of the virtual theory values can be determined. Once the procedure has been established and tested in virtual reality, and a satisfactory agreement between the virtual and real experimental data has been obtained, the procedure can be applied in toto to investigation of the empirical reality. The final result of these mathematical operations consists of confidence intervals for the measured physical values (data points) and exclusion ranges for probability distributions of the tested parameters of the physical theory.

Similar to the fact that there is no absolute space and no absolute time in physics, there is thus no absolute truth. Due to thermodynamical and quantum effects necessarily coupled to the electromagnetic interactions in the detection processes, even a fixed combination of theory parameters in top-down calculations cannot lead to any desired degree of precision and unambiguity in the measurement results. We improve our knowledge relative to the existing knowledge, which is itself in the same sense relative to its context.

The concept of knowledge appears here in two different roles: In order to be able to deduce physical measuring points from the measured positions, times and charges, physical expertise must be combined with methodological knowledge of computer science. This way a world of virtual reality is created, in which there are no unknowns. Here, the methods of inference can be optimized and confirmed. With these methods, the relevant signal is filtered out of the irrelevant background, and the resulting measured signal is assigned a physical meaning. The models of inference result from an optimized adaptation of many mathematically generated or measured examples to a target function. Though the inference models are developed in simulated world without unknown facts, as a result of these analysis processes, for physical and mathematical reasons only probability ranges of the measurements are obtained.

According to our general understanding, most of empirical reality is currently well described by theoretical physics. For many questions, however, further improvements are necessary and possible. Such theory models extending our standard understanding may be heterogeneous and characterized by the existence of approaches that may be hardly experimentally distinguishable. Because of the possible multitude of free parameters used for this purpose in theory, Monte Carlo methods are often the only way to determine

whether and in which parameter range of which theory the theoretical predictions can experimentally be tested and which parameter combinations have been constrained by a measurement. Knowledge can be here interpreted as an at least allowed or better highly preferred parameter region in the world of theories.

The model of knowledge acquisition discussed here substantially differs from Popper's critical rationalism (to which it is otherwise very much indebted) by replacing purely logical decisions with a cycle of probabilistic evolutionary evaluations. As with Popper, completely wrong models of explanation can easily be sorted out. Which probabilities one has to demand as a criterion for specific theoretical approaches not further to be pursued depends on the significance of the physical question under consideration, the customs of the physical discipline concerned, and the available resources. Methodologically, decisions based on only moderate differences in the probabilities of the explanatory approaches are not necessary. Due to the very good quality of the physical description of nature, many long used models and theories in practice have a very high probability. In this way, the classical approach to knowledge can be understood as a borderline case. And, as always, when the accustomed result is understood as a borderline case of a more general consideration, we become observers and promoters of a new phase of science.

To summarize, this article introduces the current methodology of knowledge acquisition of physics, which is based on Monte Carlo simulations and machine learning. Following Popper, but also in contrast to his approach, this new methodology is best characterized as "probabilistic rationalism". Future work in philosophy of science will have to investigate its relation to the established epistemological theories and discuss its philosophical consequences.

6 Acknowledgment

This article has been inspired by many discussions in joint teaching sessions with Brigitte Falkenburg (on philosophy of physics) and Katharina Morik (on epistemology of computer science) as well as by the long research in the DFG CRC 876. The article is further based on many years of lecturing on "Statistical Methods of Data Analysis in Physics", which was intensively discussed and improved by many members of my group deserving my cordial thank. I also thank Peter Biermann, Dominik Elsässer and Christian Spiering for very helpful remarks and discussions.

References

[1] Karl Popper, The Logic of Scientific Discovery, XXVII 513, Routledge, London [et al.] (2002)

[2] Plato. The Republic (translated by Benjamin Jowett), Lerner Publishing Group, 254, ProQuest Ebook Central, USA (2018)
https://ebookcentral.proquest.com/lib/dortmundtech/detail.action?docID=5444385.

[3] Print made by Jan Saenredam after Cornelis Cornelisz. van Haarlem published by: Hendrik Hondius I, https://www.britishmuseum.org/collection/object/P_1852-1211-120

[4] Galileo Galilei, The Assayer (translated by Stillman Drake), University of Pennsylvania Press, 151-336, Pennyslvania (2016), in The Controversy on the Comets of 1618, DOI: https://doi.org/10.9783/9781512801453-006

[5] Newton, Isaac (1726): Principia Mathematica Philosophiae Naturae, 3rd ed. Quoted after: The Principia: A New Translation by I.B.Cohen and A.Whitman, 1999. Berkeley: University of California Press.

[6] Isaac Newton, Opticks, Dover Publication Inc, 406, New York (1979)

[7] Wolfgang Wagner, Design and realisation of a new AMANDA data acquisition system with transient waveform recorders, PhD Thesis, Dortmund (2004)

[8] Brigitte Falkenburg and Wolfgang Rhode (eds.), From Ultra Rays to Astroparticles: A Historical Introduction to Astroparticle Physics, 345, Springer, Dordrecht Heidelberg New York London (2012)

[9] Brigitte Falkenburg, Pragmatic Unification, Observation and Realism in Astroparticle Physics. In: General Journal for Philosophy of Science Volume 43, Issue 2, Page 327-345 (2912)

[10] Falkenburg, Brigitte (2012): From Waves to Particles and Quantum Probabilities. In: B. Falkenburg and W. Rhode (eds.), From Ultrarays to Astroparticles. A Historical Introduction to Astroparticle Physics. Springer: Dordrecht, 265-295

[11] Katharina Morik and Claus-Rainer Rollinger, The Real Estate Agent - Modelling Users by Uncertain Reasoning, The AI Magazine, Page 44-52 (1985)

[12] J. -H. Koehne, K. Frantzen, M. Schmitz, T. Fuchs, W. Rhode, D. Chirkin, J. Becker Tjus; PROPOSAL: A tool for propagation of charged leptons, Computer Physics Communications, Volume 184, Issue 9, p. 2070-2090 (2013)

[13] Mario Dunsch, Jan Soedingrekso, Alexander Sandrock, Maximilian Meier, Thorben Menne, Wolfgang Rhode; Recent improvements for the lepton propagator PROPOSAL, Computer Physics Communications, Volume 242, p. 132-144 (2019)

[14] Jan Soedingrekso, PhD Thesis TU Dortmund, in preparation (2021)