Записки научных
семинаров ПОМИ
Том 316, 2004 г.

V. Kreinovich, A. M. Finkelstein

TOWARDS APPLYING COMPUTATIONAL COMPLEXITY TO FOUNDATIONS OF PHYSICS

ABSTRACT. In one of his early papers, D. Grigoriev analyzed the decidability and computational complexity of different physical theories. This analysis was motivated by the hope that this analysis would help physicists. In this paper, we survey several similar ideas that may be of help to physicists. We hope that further research may lead to useful physical applications.

1. INTRODUCTION

1.1. How it all started for us. Our first serious attempt to analyze the relation between computational complexity and foundations of physics was in the early 1970s, when both Dima Grigoriev and the first author were students at the Mathematics Department of St. Petersburg University. The first author knew a little bit about physics, especially about space-time geometry, and he felt that Dima's (already then) experience in decidability of the first-order theory of real numbers (and its fragments) could be very helpful in analyzing different space-time models.

We "clicked" nicely, and we wrote a joint paper in which we analyzed decidability of different space-time models. Some of our results were rather trivial: e.g., that, due to Tarski' theorem [38], the first order theory of Minkowski (flat) space-time, with the causal relation $(t, x_1, x_2, x_3) \succ (s, y_1, y_2, y_3)$ if and only if $t - s > \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2}$, is decidable. Some of our results were new (although still not very difficult to prove): that for some physically meaningful classes of space-time models, the first order theory of the causality relation is decidable, while for other physically meaningful classes, the corresponding theory is undecidable. For decidable theories, we also looked into the computational complexity of the corresponding deciding algorithms.

We submitted our paper for the student paper competition organized by the Mathematics Department and we got the Best Paper award for it. Alas, we never transformed this paper into a publication; this new paper is an attempt to publish some related ideas – in the form in which these ideas look now, 30 years later.

1.2. Why is this interesting? The first (mundane) reason. Why is it interesting to analyze the relation between computational complexity and foundations of physics? Why was it interesting to us then, and why is this research still of big interest to us?

Of course, there is a mundane reason: One of the objectives of physics is to predict the future of the world. We are not talking about the future in any grandiose sense but simply about predictions: where will the Moon be in a month (this we can do very well), what will the weather be tomorrow (this we can also do reasonably well now), where and when will the next earthquake happen (this we cannot do well yet), etc. In many such problems, the equations are well known, the initial conditions are measured with a reasonable accuracy, the main problem is that computations take so long a time that the predicted event occurs before the predicting computations end. This is why supercomputers are used in weather prediction, and this is why long-term weather prediction is not yet practically possible. Of course, from the viewpoint of a computer engineer, this is an excellent motivation to design even faster supercomputers. However, from the viewpoint of a computer scientist, before trying to build high-cost hardware that would run the existing algorithms faster, it is desirable to first check if we can design faster algorithms.

In other words, it is desirable to analyze the computational complexity of the existing physical problems – with the hope that some of these problems can be solved by much faster algorithms than now.

This is a well-known activity, relating computational complexity and working physics, many researchers in computer science and mathematical physics are working on it. What interested us more was the relation not with *working* physics but rather with *fundamental* physics. Let us explain the related possibilities in more detail.

1.3. Why is this interesting? This type of activity has led to the successes of quantum computing. Suppose that we analyze a prediction problem – like predicting geometry of space-time. If it turns out that this problem has a reasonable computational complexity, great. But what if this problem turns out to be computationally difficult? In other words, what if the future event always occurs before we can predict it?

At first glance, it may sound as if this inability to predict is bad, it limits our ability to cognize the world. However, it is possible to put a positive twist on this seemingly negative result. Namely, what such a result means is that the analyzed physical phenomenon provides us with a unique opportunity to compute the results faster than any of the existing computers – and thus, we can use this phenomenon in designing new faster computers. This positive twist can be traced to Lobachevsky, one of the fathers of non-Euclidean geometry. When Lobachevsky found out that in his non-Euclidean geometry, formulas for the volumes and areas are much more complex than in the standard Euclidean geometry, a fact that seemed to make his geometry less desirable, he immediately suggested that this complexity may be a good thing. For example, the fact that the expression for the volume is complex means that we can then compute the value of the corresponding expression by simply filling a spherical shell with water and measuring the resulting volume.

A similar positive twist on a seemingly negative result is behind the current interest in quantum computing. Indeed, for Lobachevsky geometry:

- the expressions for volumes and areas are difficult to compute;
- therefore, Lobachevsky suggested that the use of physical devices in non-Euclidean space can speed up computations of the corresponding expressions.

Similarly:

- it is very difficult to solve the quantum mechanics equations;
- it is therefore reasonable to suggest that quantum phenomena, when properly arranged, can speed up computations.

And indeed, by using a proper quantum arrangement, we can reduce the time needed to search in an unsorted n-element array from n to $O(\sqrt{n})$ [14, 15, 16], and the time needed to factor large integers from probably exponential to definitely polynomial [36]; see also [33].

1.4. Why is this interesting? Because not only can physicists help us, we can help physics as well. When we have a well-established physical theory like quantum mechanics, and this theory implies that fast prediction is impossible, then we have no other choice but to take this impossibility at face value, and design ways of making the best use of this seeming inconvenience.

In fundamental physics, however, few theories are as well established as quantum mechanics. Be it cosmology, be it quantum field theory, there are often several competing theories (or at least several competing versions of the same theory). In this situation, if one of the competing theories leads to efficient predictions while in the competing theory, prediction is impossible, then this cognizability may be a good argument in favor of the first theory.

This is what interested Dima and the first author: that by analyzing computational complexity of different physical problems, we may be able to help physicists understand the world.

This is a very ambitious desire but, as we will try to show in this paper, it is quite doable. Most of the following examples are rather mathematically trivial but in many of these cases, we gain a somewhat better understanding of physicists' intuition and/or of the relation between some physical constants.

A word of caution is in order: from the viewpoint of physics, these examples are simply a humble beginning. This is not an opportunity to brag about the results, this is rather an opportunity to present the main ideas – and thus, to hopefully encourage further research in this direction.

What we are doing is best described by the term that is widely used in engineering and applied science: "proof of concept". To avoid confusion, we should emphasize that this term has nothing to do with logical or mathematical proofs. It is most frequently used in the following context: before a granting agency allocates several million (or even billion) dollars to develop a new project (a new radiotelescope, a new particle accelerator, etc.), the agency wants to have some level of confidence that the new ideas behind this project can indeed lead to new scientific results. For example, before the granting agency pays for the efforts to connect a gravity measurements database with the seismic measurements, they would like to see, on a simplified example (with real data or even with simplified "toy" data) that such a connection can indeed lead to more accurate geophysical predictions. What we want this paper to do is to provide such a "proof of concept" that computational complexity ideas can indeed be helpful in foundations of physics.

1.5. What we are planning to do. As we have mentioned, there are two possible consequences of the collaboration between physics and computational complexity:

- on the one hand, this collaboration can help computing: namely, it can lead to a physical speed-up of computations; this is what quantum computing does;
- on the other hand, this collaboration can also help physics: namely, it may lead to a better understanding of some issues related to foundations of physics.

In this paper, we will describe ideas of both types.

This text is not so much a survey of results but rather a survey of (somewhat) related ideas. We are *not* presenting a well-developed well-connected theory. Instead, we present a kaleidoscope of loosely related ideas, if you wish, a "salad" of ideas.

Preliminary versions of (most of) these ideas have previously appeared in print, mainly as short extended abstracts in conference proceedings. This paper is our first attempt to present these related ideas together in a more detailed form.

Some of our discussion of these ideas simply involves back-of-the-envelope calculations; in some other cases, we prove some relevant results. From the mathematical viewpoints, the corresponding proofs are very simple; for a professional, once these results are formulated, their proofs are short and directly follow, in a reasonably straightforward way, from our definitions. Nevertheless, we decided to include these proofs because, in our opinion, the corresponding results provide a better insight into the discussed ideas.

From the viewpoint of potential physical applications, the fact the proofs clearly and directly follow from the definitions may even be viewed as an advantage: it will hopefully help physicists better understand the possible mathematical and physical consequences of these physics-related ideas.

1.6. Target audience. This text targets two audiences.

Our first objective is to describe our ideas for mathematicians, logicians, and theoretical computer scientists, researchers whose area of research intersects with Dima Grigoriev's. Since many such researchers are not specialists in physics, we tried to include brief explanations of the relevant physical notions in the text.

Our second objective is to describe our ideas for physicists and other scientists, especially physicists who are interested in how computational complexity can help in foundations of physics. To make out text more accessible for physicists, we tried not to make it too mathematically technical.

2. Computational Complexity and Balls in Lobachevsky Space

2.1. Lobachevsky geometry revisited. Let us start by revisiting the possibility of a positive twist on a negative result: if some physical process is difficult to compute, then we may be able to use this physical process to speed up computations. We have mentioned that this twist has been proposed by Lobachevsky himself, and that a similar twist has led to quantum computing.

In this section, we will show that a similar positive twist can also be used, in a different way, in complexity related to space-time geometry.

Specifically, a model of Lobachevsky geometry can be characterized by its *curvature* k or, alternatively, by its *characteristic distance* d_0 for which $k = 1/d_0^2$. Crudely speaking, when the distances d between the points are much smaller than this characteristic distance d_0 ($d \ll d_0$), then for computing the areas, angles, volumes, etc., we can use, with a good accuracy, formulas from the Euclidean geometry. To be more precise, for each of the corresponding geometric expressions e, the corresponding Lobachevsky formulas have the form $e_{\rm Lob} = e_{\rm Eucl} + O(d/d_0)$, where $e_{\rm Eucl}$ is the corresponding Euclidean expression. When the distances approach d_0, then the results of applying the Lobachevsky-geometry formulas can become very different from the results of applying more traditional Euclidean-geometry expressions.

In particular, in contrast to Euclidean geometry, where the volume $V(R)$ of a ball of a radius R grows as R^3, in Lobachevsky geometry, this volume increases exponentially with R: $V(R) \approx K^R$ for $K = \exp(1/d_0) > 1$.

Let us show that, because of this exponential growth, in Lobachevsky space-time, we can, in some sense, solve propositional satisfiability in polynomial time.

The main idea behind this solution is as follows: for every n, we:

- place exponentially many (2^n) processors in a ball of radius $R \approx n \cdot \log_2(K)$ (for which $V(R) \approx K^R = 2^n$);
- let each of these processors test a different Boolean vector, and
- let the one who found a satisfying Boolean vector send the result back to us.

Checking a single Boolean vector is fast, so, since the largest distance d is $\approx n$, the communication time d/c – where c is the speed of light – is also polynomial in n.

This idea does not directly lead to a polynomial-time solution because to implement this idea, we need to actually place all the processors within the ball. If we place all the processors one by one, then, of course, we will still need an exponential time to place all these processors.

So, if we are only interested in solving a single satisfiability problem with n variables, there seems to be no advantage in using this scheme.

However, in real life, we are interested in many possible satisfiability problems of size n (or, alternatively, problems from other NP classes). In this case, if we use the above Lobachevsky-space scheme, we only need to spend exponential time for solving the very first problem: after that, we will have 2^n processors already placed in a ball of radius $O(n)$, and we will be able to use this ready-made processor configurations to solve all other problems of the same size.

It is therefore (somewhat) reasonable to assume that we have as many processors as needed already placed there.

It is also potentially possible to place all the necessary exponential number of processors in polynomial time. For that, we must assume that – like in our physical space – in terms of large-scale distributions, matter is more or less uniformly distributed in the Lobachevsky space. So, within the ball of radius $O(n)$, there are already exponentially many planets on which we can place the processors. To place the processors, we send two specially programmed self-propelling self-replicating robots to the two nearest planets. Each of these robots, after landing on a planet, builds a small plant that produces three things: a processor to stay on this planet, and two more self-replicating robots of exactly the same type. Each of these new robots, in turn, will launch themselves to the nearest not yet exploited planets, etc. With this scheme, on each stage, the number of robots doubles, hence the number of processors doubles; so, after n stages, we get 2^n processors. Each stage consists of building a plant – the time for that does not depend on n – and travel between the planets which, as we have mentioned, is limited by $O(n)$. Thus, the overall time for all n steps is bounded by $n \cdot O(n) = O(n^2)$ – i.e., we can fill the nearby space with 2^n processors in polynomial time.

It is worth mentioning that if we make the robots move smartly, always in such a way that the distance from the original planet does not decrease, then the overall travel time will only grow as $O(n)$.

In Euclidean space-time, we can collect the same exponential number of processors but these processors will then require an exponential distance – hence exponentially growing communication time.

This scheme, originally proposed in [23, 32], is described in more detail in [18]. Of course, the real physical space-time is more complex than Lobachevsky space; however, a similar speed-up can be achieved for more physically reasonable space-times as well [20, 22, 32].

2.2. Beyond (deterministic) space-time models, towards (probabilistic) quantum physics. In the Introduction, we mentioned physics but all we have seriously discussed so far was deterministic space-time models: research into their decidability and their use in computations.

In the following text, we will venture into other areas of physics as well. It is worth mentioning that, according to modern physics, the world is not deterministic: even if knew the exact initial state of the world, we would still not be able to exactly predict the results of all future measurements – only the *probabilities* of different results. This probabilistic character of modern physics comes from the *quantum effects*: according to modern physics, the world is described by the equations of quantum mechanics, and these equations only allow probabilistic predictions. From this viewpoint, deterministic physical models are only approximations to the actual physical behavior.

Quantum mechanics was discovered when researchers started studying microscopic objects, the size of atoms and molecules. Usually, the larger the system, the smaller the corresponding quantum effects. So, when we study large-scale effects like cosmology or space-time models, we can reasonably use deterministic models. However, if we want to also describe smaller-scale objects and processes, we must take the nature's probabilistic character into consideration.

Because of this, in the following text, we will get deeply involved with probabilities, and we will also get deeply involved with quantum mechanics – as a method of describing these probabilities.

3. Computational Complexity Ideas Can Help in Explaining Physics: Proof of Concept

Our objective is to show that the computational complexity ideas can indeed be helpful in foundations of physics. Let us start with an example of the so-called "Dirac's large numbers"; in this example, no serious mathematics will be needed, only back-of-the envelope calculations. Before we start our explanations, let us first explain what are Dirac's large numbers.

3.1. There are many constants in physics. In physics, there are many constants such as the speed of light, the charge of the electron, etc. Most of these constants are *dimensional* in the sense that their numerical value depends on the choice of the measuring units for the corresponding physical quantities. For example, if we express the numerical value of the speed of light c in miles per second, we get a different numerical value than when we express it in kilometers per second. (In theoretical physics, it is often convenient to select a special system of units if which $c = 1$.)

3.2. Fundamental dimensionless constants. Some physical constants are *dimensionless* in the sense that they are independent of the choice of units. A simple example of such a constant is a ratio between the masses of a neutron and a proton. The values of most of these dimensionless constants can be derived from the corresponding physical theory. However, for some constants, we know of no such derivation, so these constants are *fundamental*.

3.3. Most dimensionless constants are not too large and not too small. The values of most fundamental dimensionless constants are usually close to 1. This fact is the reason why engineers and physicists can safely estimate and neglect, e.g., quadratic (or, in general, higher order terms) in asymptotic expansions, even though no accurate estimates on the coefficients on these terms is known [11]. In particular, such methods are used in quantum field theory, where we add up several first Feynman diagrams [2]; in celestial mechanics [37], etc.

However, there are few very large and very small ones. In 1937, P. A. M. Dirac, one of the founding fathers of quantum field theory, discovered an interesting empirical relation between such unusual constants [8, 9].

3.4. An example of a very large fundamental constant. One such fundamental dimensionless constant is related to the lifetime T of the Universe ($T \approx 10^{10}$ years; see, e.g., [11, 31, 41]). By definition, the value T is the largest physically possible time interval. To transform it into a dimensionless constant, let us divide T by the smallest possible time interval Δt. The smallest possible time is the time when we pass through the smallest possible object with the largest possible speed. The largest possible speed is the speed of light c, the smallest possible object is an elementary particle. Which of the elementary particles has the smallest size?

In Newtonian physics, particles of smaller mass m have smaller sizes but in quantum physics, the situation is different: An elementary particle does not have any components, so it is, in some sense, a point particle. However, due to Heisenberg's uncertainty principle between energy E and time t, $\Delta E \cdot \Delta t \geqslant \hbar$ (where $\hbar$ is Planck's constant), we do not see it as a point particle: the accuracy Δt with which we can locate the particle in time cannot be smaller than $\Delta t \approx \hbar/\Delta E$. Due to this uncertainty, we are not sure whether, at a given spatial location, the particle in question is present at a given moment of time. If the particle is not present, then the overall energy in this location is 0: $E = 0$. If the particle is present, then we have the energy of this particle at this point, i.e., $E = mc^2$. Thus, the uncertainty ΔE with which we know the energy is equal to mc^2. Hence, the resulting uncertainty Δt cannot be smaller than $\hbar/E = \hbar/(mc^2)$.

Comment. This is a very qualitative explanation of this inequality. For a more detailed physical explanation, see, e.g., [11].

Because of the above inequality, the smallest size particle is the one with the largest mass. Among independent stable particles – photon, electron, proton, etc. – proton has the largest mass and hence, the smallest possible Δt.

If we divide T by proton's Δt, we get a dimensionless constant $\approx 10^{40}$. There is no good physical explanation for this constant.

3.5. Dirac's relation between fundamental physical constants. Dirac noticed that this constant $\approx 10^{40}$ is unexpectedly related to another dimensionless constant: the fine structure constant $\alpha \approx 1/137$. This constant occurs in quantum electrodynamics; crudely speaking, the largest possible size of an atom is $1/\alpha$. Dirac noticed that $10^{40} \approx 2^{1/\alpha}$. Of course, this is not an exact equality but, on the other hand, we do not even know T well enough: it can be 10 billion years, it can be 20. Within the accuracy with which we know T, this coincidence was, in Dirac's viewpoint, very impressive.

3.6. Why? Since Dirac's 1937 paper, physicists have tried to explain this empirical relation. For example, Werner Heisenberg, one of the founding fathers of quantum mechanics, tried to formulate a unified quantum theory in such a way that this unified theory would explain this empirical relation (and other Dirac's relations mentioned later, in Section 4); see, e.g., [17]. Alas, these efforts did not bring an understanding of this empirical relation.

In his 1985 book [12], a famous Nobelist Richard Feynman wrote that the value $1/\alpha$ "has become a mystery ever since it was discovered more than fifty years ago, and all good theoretical physicists put this number up on their wall and worry about it. Immediately you would like to know where this number for a coupling comes from. Nobody knows. It is one of the *greatest* damn mysteries of physics: a *magic* number that comes to us with no understanding by man" (we borrowed this citation from a recent review [3], where a detailed history is also provided.)

In the late 1950s, Robert Dicke [7] showed that in a Universe in which some other Dirac's empirical relations do not hold, the carbon-based life may not be possible; this observation started the development of the so-called *Anthropic Principle* (see, e.g., [1, 31, 35]) according to which the relation between physical constants comes from the observed fact that there is life in our Universe.

We will show that simple cognizability (= computational complexity) arguments can explain this relation.

Our objective is not so much to proclaim one more explanation of the relation between the constants but rather to claim that the ideas of computational complexity can indeed lead to an explanation of this relation. We therefore hope that physicists will start looking for more serious quantitative explanations of this phenomena and convert our first-approximation idea into a true physical explanation.

3.7. Quantum mechanics: a brief overview of the main ideas. In order to describe our back-of-the-envelope explanations, we need to describe, to the readers who may not be familiar with quantum physics, what are the main ideas behind quantum descriptions of nature. Readers who have basic knowledge of quantum physics can skip this subsection and go directly to the explanations.

In Newtonian (pre-quantum) mechanics, to describe the state of a single pointwise particle at a given moment of time t, we must know the location $\vec{x}$ and the velocity $\vec{v}$ of the particle at the moment t. In three-dimensional space, we therefore need three real-valued parameters to describe the location $\vec{x}$ and three more to describe the velocity $\vec{v}$ – the total of six parameters.

Real-life measurements are never 100% accurate. As a result, due to the measurement errors, the measured values may be slightly different from the actual values of coordinate and velocity. When the measured value $\widetilde{x}$ of the first coordinate is, say, 1.0, and the measurement error $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x$ is described by a Gaussian distribution with 0 mean and standard deviation 0.1, then the only information that we have about the actual (unknown) value x of the first coordinate is that x is normally distributed with the mean 1 and standard deviation 0.1. In other case, we may end up with different probability distributions for the corresponding actual value. This probability distribution is usually characterized by the corresponding probability density $\rho(x)$.

In Newtonian mechanics, in principle, it is possible to measure both coordinates and velocity with higher and higher accuracy, so that we can get more and more accurate approximations to $\vec{x}$ and $\vec{v}$. Therefore, in Newtonian mechanics, it makes perfect sense to talk about the actual state of a particle.

To describe the state of several (n) particles at a given moment of time, we must describe the coordinates $\vec{x}_i$ and the velocity $\vec{v}_i$ of each of these n particles; so, overall, we need $6n$ parameters.

For each of these particles, after each measurement, we have a probability distribution $\rho_i(\vec{x}_i)$ that describes the distribution of the possible values of the coordinates $\vec{x}_i$. If the particles are independent and the corresponding measurements are independent, then the coordinates $\vec{x}_1, \ldots, \vec{x}_n$ of these particles are independent random variables, and so, the joint probability distribution $\rho(\vec{x}_1, \ldots, \vec{x}_n)$ of the coordinates of all n particles is equal to the product of the corresponding probability densities (see, e.g., [40]):

$$\rho(\vec{x}_1, \ldots, \vec{x}_n) = \rho_1(\vec{x}_1) \cdot \ldots \cdot \rho_n(\vec{x}_n).$$

If we also take velocities into consideration, then we get a similar probability density:

$$\rho(\vec{x}_1, \ldots, \vec{x}_n, \vec{v}_1, \ldots, \vec{v}_n) = \rho_1(\vec{x}_1, \vec{v}_1) \cdot \ldots \cdot \rho_n(\vec{x}_n, \vec{v}_n).$$

When the particles are not independent, i.e., when they interact, then there are additional restrictions that make distributions dependent as well: e.g., that the overall energy $\frac{1}{2} \cdot \sum m_i \cdot \vec{v}_i^2 + V(\vec{x}_1, \ldots, \vec{x}_n)$ should be exactly equal to its original value. In this case, we have a generic (not necessarily independent) probability distribution $\rho(\vec{x}_1, \ldots, \vec{x}_n, \vec{v}_1, \ldots, \vec{v}_n)$.

In Newtonian mechanics, in principle, for any given accuracy $\varepsilon > 0$, we can get a description of the coordinates $\vec{x}_1, \ldots, \vec{x}_n$ and velocities $\vec{v}_1, \ldots, \vec{v}_n$ of all n particles with this accuracy. In quantum mechanics, as we have mentioned, there is the Heisenberg's uncertainty principle, according to which it is not possible to measure both the coordinate and the velocity of a given particle with a given accuracy. As a result, the pre-quantum description of the particle's state by a pair $(\vec{x}, \vec{v})$ no longer has any physical meaning. Instead, to describe a quantum state, we must describe the corresponding probability distribution.

Due to Heisenberg's principle, we cannot exactly measure the coordinates and the velocities: once we get the exact values of the coordinates, the value of the velocity becomes undetermined. By measuring the coordinates several times, we can get the probability distribution $\rho(\vec{x}_1, \ldots, \vec{x}_n)$.

This probability distribution only describes the location of the particles but not their velocities. It turns out that we can describe the velocity information as well if instead of the probability density

$\rho(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n)$, we consider a complex-valued function $\Psi(\vec{x}_1, \ldots, \vec{x}_n)$ for which $|\Psi(\vec{x}_1, \ldots, \vec{x}_n)|^2 = \rho(\vec{x}_1, \ldots, \vec{x}_n)$. Here, the absolute value of the complex number $\Psi(\vec{x}_1, \ldots, \vec{x}_n)$ characterizes the particle's location, while the phase of this complex number keeps information about the possible values of velocity. The function $\Psi(\vec{x}_1, \ldots, \vec{x}_n)$ is called a *wave function*.

Similar to the Newtonian mechanics, when n particles are independent, then the wave function can be represented as a product of the corresponding 1-particle wave functions: $\Psi(\vec{x}_1, \ldots, \vec{x}_n) = \Psi_1(\vec{x}_1) \cdot \ldots \cdot \Psi_n(\vec{x}_n)$. Similarly, e.g., if particles $1, \ldots, m$ and particles $m+1, \ldots, n$ form two clusters A and B that do not interact with each other, then $\Psi(\vec{x}_1, \ldots, \vec{x}_n) = \Psi_A(\vec{x}_1, \ldots, \vec{x}_m) \cdot \Psi_B(\vec{x}_{m+1}, \ldots, \vec{x}_n)$. If all n particles actively interact with each other, then we need a general function of n coordinate variables.

According to quantum mechanics, the dynamics of an n-particle quantum system is described by Schrödinger's equation $\mathrm{i}\,\hbar \frac{\partial \Psi}{\partial t} = H\Psi$, where H is an operator depending on the exact interaction between the particles.

3.8. Our back-of-the-envelope explanation. In general, to completely describe the state of a quantum particle means to describe the probabilities $\rho(\vec{x}\,) = |\Psi(\vec{x}\,)|$ corresponding to different coordinates $\vec{x}$. Similarly, to completely describe a state of an n-particle cluster, we must know, for each combination of coordinates $(\vec{x}_1, \ldots, \vec{x}_n)$, the corresponding value $\rho(\vec{x}_1, \ldots, \vec{x}_n) = |\Psi(\vec{x}_1, \ldots, \vec{x}_n)|^2$.

It is reasonable to assume that the world is cognizable, i.e., informally, that we must be able to predict at least something for such n-particle clusters. From this viewpoint, what is the largest size n of the atom, i.e., what is the largest size of the cluster of actively interacting particles?

When we say that we must be able to predict at least *something* about the location of each of n particles, we do not mean that we should be able to predict the *complete* information about the locations of all n particles. What we would like to do is to predict at least something about each of these particles.

In other words, for each particle, we may not be able to predict its exact location but we should at least be able to have some information about this location. The smallest possible amount of information corresponds to the case when we know exactly one bit of information, i.e., an answer to a "yes"-"no" question. In other words, the smallest possible amount of information corresponds to the case when we ask one question about the location and get the answer "yes" or "no". For example, this question may be: "are we 1 mm close to a given point P?".

To describe such a situation, we must select, for each particle i, the set S_i^+ of all the locations in which the answer is "yes"; then, the set S_i^- of all the locations in which the answer is "no" is the complement to this set S_i^+.

In this case, for a single particle, we must know two probabilities: the probability $\int_{\vec{x}\in S^+} \rho(\vec{x})\,\mathrm{d}\vec{x}$ that the particle is located in the set S^+, and the probability $\int_{\vec{x}\in S^-} \rho(\vec{x})\,\mathrm{d}\vec{x}$ that the particle is located in the set S^-. These probabilities correspond to the case when instead of knowing the exact values of the probability density $\rho(\vec{x})$ for all $\vec{x}$, we know only two integral values of this probability density corresponding to $\vec{x} \in S^+$ and $\vec{x} \in S^-$.

Similarly, for an n-particle cluster, if we want to describe, for each particle i, whether its coordinates $\vec{x}_i$ belong to the set S_i^+ or to the set S_i^-, we must describe, for each possible combination of sets $S_i^\pm$, what is the corresponding integral probability. For n particles, there are exactly 2^n possible combinations $(S_1^\pm, \ldots, S_n^\pm)$, so we must know at least 2^n different probabilities.

In the case of several clusters that are independent from each other, we can use the fact that the corresponding probabilities are equal to the products of the probabilities corresponding to different clusters, and thus reduce the required overall number of probabilities. However, if all the particles form a single cluster, with strong interaction and thus, no independence, we have to consider all 2^n possible values.

To describe 2^n different values of probability, we must consider at least 2^n different values of the corresponding wave function $\Psi(\vec{x}_1, \ldots, \vec{x}_n)$.

Our prediction algorithm must handle each value at least once, so it is reasonable to conclude that this algorithm should require at least 2^n computational steps[1]. The shortest possible time of a computation step is the shortest possible time interval Δt. Thus, during the entire history of the Universe, we can perform no more than $T/\Delta t$ computational steps. The largest possible atom is thus the one for which we need this largest number of steps, i.e., for which $2^n \approx T/\Delta t$. This is exactly Dirac's relation.

Comments. In the above text, we considered the "minimal case" of, in effect, two possible locations of each particle. We have shown that in

[1] See comment below on the possibility of using quantum computers.

this case, we can make predictions only about clusters of size $\leqslant 1/\alpha$. Alternatively, we can consider the cases when more information can be predicted about each particle from a cluster. For example, if for some k, we allow k different alternative pieces of information about each particle, then we need k probabilities to store the information about each particle and k^n probabilities to store the information about the n-particle cluster. To be able to process this information in time $\leqslant T\Delta t$, we must therefore require that $k^n \leqslant T/\Delta t$. Thus, the size n of the cluster cannot exceed $\log_k(T/\Delta t) = \log(T/\Delta t)/\log(k)$.

The larger k, the smaller this size; the largest possible size corresponds to the smallest possible k. So, if we want to be able to predict, say, 3 bits of information about a location of each particle in a cluster, then we can only do it with clusters of size $\ll 1/\alpha$. The larger the cluster, the fewer information we can get about each particle in this cluster. When the cluster size reaches $1/\alpha$, we can get exactly 1 bit of information about the location of each particle. For clusters of larger size, we have even fewer bits of information about the locations of different particles – which means that for some of the particles, we will have no information about their location at all. Thus, the largest clusters for which we can predict some information about the (location of) each particle from a cluster is indeed of size $\approx 1/\alpha$.

Readers should be warned that this is a “first approximation”, back-of-the-envelope calculations. In reality, the situation is more complicated:

- On the one hand, we may speed up computations if we use quantum or parallel computation, so in time $T/\Delta t$, we can, in effect, make more than $T/\Delta t$ computational steps.
- On the other hand, if we want to predict whether the location $\vec{x}_i$ of i-th particle in within a set S_i or not, it is not sufficient to just know whether $\vec{x}_i \in S_i$ in the initial state. We usually need more information about the original location of a particle to predict its location at a future moment of time. As a result, we need to process much more than 2^n units of data.

Our back-of-the-envelope calculations provide a qualitative explanation of Dirac's relation. In order to lift this qualitative explanation to a more accurate and reliable physical level, we must know, e.g., how different uncertainties in the initial state of a system propagate in time. This propagation, in its turn, may be different for different alternative equations that describe the evolution of a cluster. In other words, if we provide a more detailed analysis, we may have different cluster sizes depending on what type of interactions we consider between the particles that form this cluster.

4. The Notion of the Smallest Probability

4.1. Other Dirac's relations. Dirac has discovered several other relations between dimensionless constants. These relations can also be justified in a similar way but this justification requires a deeper physical analysis than simple back-of-the-envelope calculations from the previous section.

Most of the analysis in this section is still on the physical level of rigor. In the next section, we show how many of these results can be described in precise mathematical terms.

Let us start our analysis by formulating an important difference between the physical intuition and the existing mathematical formalisms.

4.2. Physicists assume that initial conditions and values of parameters are not abnormal. To a mathematician, the main contents of a physical theory is the equations. The fact that the theory is formulated in terms of well-defined mathematical equations means that the actual field must satisfy these equations. However, this fact does *not* mean that *every* solution of these equations has a physical sense. Let us give three examples:

Example 1. At any temperature greater than absolute zero, particles are randomly moving. It is theoretically possible that all the particles start moving in one direction, and, as a result, a person starts lifting up into the air. The probability of this event is small (but positive), so, from the purely mathematical viewpoint, we can say that this event is possible but highly unprobable. However, the physicists say plainly that such an abnormal event is *impossible* (see, e.g., [10]).

Example 2. Another example from statistical physics: Suppose that we have a two-chamber camera. The left chamber is empty, the right one has gas in it. If we open the door between the chambers, then the gas would spread evenly between the two chambers. It is theoretically possible (under appropriately chosen initial conditions) that the gas that was initially evenly distributed would concentrate in one camera. However, physicists believe this abnormal event to be impossible. This is an example of a "micro-reversible" process: on the atomic level, all equations are invariant with respect to changing the order of time flow ($t \to -t$). So, if we have a process that goes from state A to state B, then, if while at

B, we revert all the velocities of all the atoms, we will get a process that goes from B to A.

However, in real life, many processes are clearly irreversible: an explosion can shatter a statue but it is hard to imagine an inverse process: an implosion that glues together shattered pieces into a statue. Boltzmann himself, the 19th century author of statistical physics, explicitly stated that such inverse processes "may be regarded as impossible, even though from the viewpoint of probability theory that outcome is only extremely improbable, not impossible." [4].

Example 3. If we toss a fair coin 100 times in a row, and get heads all the time, then a person who is knowledgeable in probability would say that it is possible – since the probability is still positive. On the other hand, a physicist (or any person who uses common sense reasoning) would say that the coin is not fair – because if it is was a fair coin, then this abnormal event would be impossible.

In all these cases, physicists (implicitly or explicitly) require that the actual values of the physical quantities must not only satisfy the equations but they must also satisfy the additional condition: that the initial conditions should *not* be *abnormal*.

Comment. In all these examples, a usual mathematician's response to physicists' calling some low-probability events "impossible", is just to say that the physicists use imprecise language.

It is indeed true that the physicists use imprecise language, and it is also true that in the vast majority of practical applications, a usual probabilistic interpretation of this language perfectly well describes the intended physicists' meaning. In other words, the probability language is perfectly OK for most physical applications.

However, there are some situations when the physicists' intuition seem to differ from the results of applying traditional probability techniques:

- From the probability theory viewpoint, there is no fundamental difference between such low-probability events as a person winning a lottery and the same person being lifted up into the air by the Brownian motion. If a person plays the lottery again and again, then – provided that this person lives for millions of years – he will eventually win. Similarly, if a person stands still every morning, then – provided that this person lives long enough – this person will fly up into the air.

- On the other hand, from the physicist viewpoint, there is a drastic difference between these two low-probability events: yes, a person will win a lottery but no, a person will never lift up into the air no matter how many times this person stands still.

We have just mentioned that the traditional mathematical approach is to treat this difference of opinion as simply caused by the imprecision of the physicists' language. What we plan to show is that if we take this difference more seriously and develop a new formalism that more accurately captures the physicists' reasoning, then we may end up with results and directions that are, in our opinion, of potential interest to foundations of physics. In other words, what we plan to show is that if we continue to use the traditional probability approach, it is perfectly OK but if we try to formalize the physicists' opinion more closely, we may sometimes get even better results.

4.3. A natural formalization of this idea. The above-mentioned property of being "not abnormal" ("typical") has a natural formalization: if a probability of an event is small enough, i.e., $\leqslant p_0$ for some very small p_0, then this event cannot happen.

In other words, there exists the "smallest possible probability" p_0 such that:

- if the computed probability p of some event is larger than p_0, then this event can occur, while
- if the computed probability p is $\leqslant p_0$, the event cannot occur.

For example, the probability that a fair coin falls heads 100 times in a row is 2^{-100}, so, if the threshold probability p_0 satisfies the inequality $p_0 \geqslant 2^{-100}$, then we will be able to conclude that such an event is impossible.

In the next section, we will see that this formalization is not perfect (and how computational complexity can help) but as of now, let us describe the physical consequences of this assumption and, based on the known physics, estimate the value of this "smallest probability" p_0.

In order to estimate this "smallest probability" p_0, we will consider several situations in which the existence of such smallest probability will lead to other physical consequences like the smallest spatial accuracy with which we can detect a particle, the longest time during which we can observe the particle, etc. Each of these consequences is similar to what is actually observable in quantum theory. By comparing estimates coming from the existence of the smallest probability and the estimates coming from the more traditional quantum theory, we are therefore able to estimate the corresponding value of p_0.

In other words, to get first-approximation estimates for p_0, we do what physicists often do to get first-approximation estimates: we perform some reasonable speculations and manipulations with the formulas of quantum theory.

4.4. First estimation of p_0: based on the smallest physically meaningful spatial size. We have already mentioned, in the previous section, that the range of particle's time location is $\Delta t \approx \hbar/(mc^2)$, so the range of the particle's spatial location is $\Delta x = c \cdot \Delta t \approx \hbar/(mc)$.

As we have also mentioned, in quantum physics, an elementary particle is not located at an exact point with 100% probability but it is, nevertheless, a point-wise particle. In other words, the fact that we cannot locate a particle with an accuracy better than Δx does not mean that smaller spatial sizes $\varepsilon < \Delta x$ do not have physical sense. In principle, we can measure the particle's location with a higher accuracy but this higher-accuracy value does not mean that the particle is located at this particular point within the Δx-size range. The particle is characterized by a probability distribution, not by an exact location value, and repeated measurements will lead to different values within the particle's spatial range. For every zone within the Δx-size range, there is a probability that the coordinates measurement will result in coordinates within this particular zone.

According to quantum mechanics, the probability density of a particle distribution is equal to $|\Psi(\vec{x}\,)|^2$. Thus, the probability p to find a particle in a zone of linear size ε around a point $\vec{x}$ is $\approx |\Psi(\vec{x}\,)|^2 \cdot \varepsilon^3$.

If we assume that events with probability $\leqslant p_0$ cannot happen, then we get the following phenomenon: When we decrease ε, at first, we get more and more accurate location of the particle. However, when the spatial size ε gets very small, the probability p becomes smaller than the "smallest probability" p_0, so we will not see anything at all [21].

Comment. This is similar to what Chris Kelvin, the hero of Lem's "Solaris" [29], finds when he analyzes the ghost creature resembling his ex-wife: when he increases the amplification of the electronic microscope, he sees more and more details – until he sees nothing at all. The difference between Lem's science fiction and the actual physical picture is that in Lem's vision, this phenomenon is only true for ghost creatures, while in reality, it is true for all the universe's matter.

For each value of p_0, we can estimate the smallest spatial size ε_0 for which we can still see something, i.e., the smallest spatial size that still makes physical sense. In the first approximation, it is reasonable to assume that the probability distribution $\rho(\vec{x}\,) = |\Psi(\vec{x})|^2$ is isotropic Gaussian:

$$\rho(\vec{x}\,) = \frac{1}{(\sqrt{2\pi})^3 \cdot \sigma^3} \cdot \exp\left(-\frac{(\vec{x} - \vec{x}_0)^2}{2\sigma^2}\right),$$

with the standard deviation σ equal to the particle's spatial range $\Delta x = \hbar/(mc)$. The largest possible value of the probability density $\rho(\vec{x}\,) = |\Psi(\vec{x}\,)|^2$ is attained at the central point $\vec{x} = \vec{x}_0$ and is equal to $\approx \sigma^{-3} = (\Delta x)^{-3}$. Thus, when $(\Delta x)^{-3} \cdot \varepsilon^3 \leqslant p_0$, we conclude that $p \leqslant p_0$ for all points $\vec{x}$ – hence, within this spatial size, we will not be able to see anything at all.

On the other hand, if $(\Delta x)^{-3} \cdot \varepsilon^3 > p_0$, then the probability to observe the particle in the ε-vicinity of its center x_0 is greater than the smallest probability p_0. So, for such values ε, we can use the measuring instrument with a spatial resolution ε and still observe the particle.

So, the smallest possible spatial size ε_0 that makes physical sense is determined by the equation $(\Delta x)^{-3} \cdot \varepsilon_0^3 = p_0$. Due to this equation, this size is equal to $\varepsilon_0 = p_0^{1/3} \cdot \Delta x$.

4.5. What is the smallest physically possible probability p_0? We have shown that for every value p_0, we can find the smallest possible spatial resolution with which we can detect a particle: $\varepsilon_0 = p_0^{1/3} \cdot \Delta x$.

On the other hand, in physics, it is known that there is a length ε_0 beyond which (due to quantum effects in geometry) we cannot localize anything (see, e.g., [31]), the so-called Planck length $\varepsilon_0 = \sqrt{\hbar \cdot \gamma / c^3} \approx 10^{-33}$ cn, where γ is the gravitational constant. Equating the two expressions for ε_0 and taking into consideration that $\Delta x = \hbar/(mc) \approx 10^{-13}$ cm, we conclude that $p_0 \approx 10^{-60}$.

4.6. Lifetime of a free particle – and another way of estimating p_0. According to Schrödinger's equation (see, e.g., [11]), the Gaussian free particle is not a static solution: as time grows, the state continues to be Gaussian but it expands, so that at a time t, its range is $\Delta x(t) \approx \sqrt{\hbar t/m} = \sqrt{\Delta x \cdot (ct)}$, where $\Delta x = \hbar/(mc)$ denotes the standard spatial range of the particle.

The probability to notice the presence of the particle within its standard range of linear size Δx is proportional to $\rho(\vec{x}\,) \cdot \Delta x^3$; similar to localization case, we can conclude that the largest possible value of $\rho(\vec{x}\,)$ is $\approx (\Delta x(t))^{-3}$. So, when the corresponding probability $\rho(\vec{x}\,) \cdot \Delta x^3 \approx$

$(\Delta x(t))^{-3} \cdot \Delta x^3$ is larger than or equal to the smallest probability p_0, we can still find the particle in its normal-size range. On the other hand, when $(\Delta x(t))^{-3} \cdot \Delta x^3 < p_0$, we will not be able to find the particle in its normal-size range: the particle kind of disappears.

The larger the time t, the larger the radius $\Delta x(t)$ and the smaller the corresponding probability $(\Delta x(t))^{-3} \cdot \Delta x^3$. When $t \to \infty$, this probability tends to 0 and thus, eventually becomes smaller than the smallest probability p_0. Once the probability to find the particle in its normal size range becomes smaller than the smallest possible probability p_0, then the particle disappears.

In other words, in this scheme, free particles have a finite lifetime. For any given value p_0, we can estimate the lifetime T of a free particle as the largest value T for which the corresponding probability is still $\geqslant p_0$, i.e., as the value for which $(\Delta x(T))^{-3} \cdot \Delta x^3 = p_0$. Substituting the expression $\Delta x(T) = \sqrt{\Delta x \cdot (cT)}$ into this equation, we conclude that

$$\frac{\Delta x^3}{(\Delta x \cdot (cT))^{3/2}} = \frac{\Delta x^{3/2}}{(cT)^{3/2}} = p_0,$$

i.e., that $cT/\Delta x = T/\Delta t = p_0^{-2/3}$.

From the physical viewpoint, it is reasonable to assume that the lifetime T of the free particle is equal to the lifetime of the Universe. For the lifetime of the Universe, the ratio $T/\Delta t$ is one of Dirac's large numbers, and it is known to be equal to 10^{40}. Thus, from $10^{40} = p_0^{-2/3}$, we conclude that $p_0 \approx 10^{-60}$ – the same value as before.

4.7. Number of particles in the Universe, one more Dirac's relation, and the third estimate of p_0. Let us now show that another consequence of our assumption – that there exists the smallest probability – is that the overall number of particles in the Universe is limited.

In the previous analysis, we considered individual particles. In quantum physics, specifically in quantum field theory, it is known that the real state of the world may not have a fixed number of particles, we may have different number of particles with different probabilities. As a result, when we combine many particles together, often, we do not get a state with a fixed number k of particles, we get a state in which we may get different values k with different probabilities $p(k)$.

We have assumed that events with a low probability ($\leqslant p_0$) cannot occur while events with probability $> p_0$ can occur. Thus, if $p(k) \leqslant p_0$ for all k, then we simply cannot count the number of particles because no outcome of this counting is physically possible. We will show that this indeed happens when we combine too many particles together and thus, that there is a physical limit on the number of particles that we can actually count.

In order to explain how this happens let us first go back to the state of a single particle. From the viewpoint of the above "smallest probability" assumption, what we consider physically as a single particle may be, from the mathematical viewpoint, a multi-particle state, in which the probability of measuring the number of particles as 0 (or 2, 3, etc.), is $\leqslant p_0$. How can we describe this situation in exact terms? In other words, how can we describe the probability $p(k)$ of this seemingly single particle being actually a k-particle state?

Our main assumption is that events with probability $\leqslant p_0$ cannot happen. Thus, the fact that we observe a particle as a 1-particle state means that the (unknown) probabilities $p(0)$, $p(2)$, $p(3)$, $\ldots$, all belong to the interval $[0, p_0]$ (of impossibly low probabilities). We do not know the exact values of $p(k)$ within this intervals, so it may seem natural to assume that these (unknown) values are uniformly distributed within the corresponding intervals. In this case, the average low probability is equal to $p_0/2$.

The problem with this randomness assumption is that in this case, due to the central limit theorem, the overall probability $p(0) + p(2) + p(3) + \ldots + p(k) + \ldots$ of the particle to be in one of these states grows with k as $k \cdot (p_0/2)$ – and therefore eventually exceeds 1. To avoid this absurd situation, we must install a bound n_0 on how many particle can actually be in this one-particle state. As a result, we arrive at the following model of a single particle: for some n_0, the probability $p(k)$ is uniformly distributed within $[0, p_0]$ for $k \leqslant n_0$, and $p(k) = 0$ for $k > n_0$.

The value n_0 can be determined by the fact that in this state, the expected number of particles should be approximately 1: $\bar{n} = \sum\limits_{k=0}^{n_0} k \cdot p(k) \approx 1$. Since the expected value of $p(k)$ is $p_0/2$, this expected value $\bar{n}$ of n is equal to $\approx p_0/2 \cdot \sum\limits_{k=0}^{n_0} k \approx p_0 \cdot n_0^2/4$. Hence, from $\bar{n} \approx 1$, we conclude that $n_0 \approx p_0^{-1/2}$.

Thus, the standard deviation σ of the number of particles in this state is equal to

$$\sigma \approx \sqrt{\sum_{k=1}^{n_0} k^2 \cdot p(k)} \approx \sqrt{n_0^3 \cdot p_0} \approx p_0^{-1/4}.$$

If we combine N independent particles, then the (random) overall number of particles in the resulting state is equal to the sum of N independent random variables that describe the numbers of particles in the composed states. Due to the central limit theorem, for large N, the distribution of the number of particles $p_N(k)$ in the resulting composite state is Gaussian, with the average and the standard deviation of the composite state equal to $N \cdot \bar{n}$ and $\sigma_N = \sigma \cdot \sqrt{N}$:

$$p_N(k) = \frac{1}{\sqrt{2\pi} \cdot \sigma_N} \cdot \exp\left(-\frac{(k - N \cdot \bar{n})^2}{2\sigma_N^2}\right).$$

The largest of these probability values is attained when $k \approx N \cdot \bar{n}$, and it is equal to $\approx 1/\sigma_N$. Thus, while $1/\sigma_N > p_0$ and we count the particles (i.e., we measure the number of particles in this configuration), then one of the probabilities $p_N(k)$ is larger than the threshold p_0, and thus, one counting example is possible. On the other hands, when $1/\sigma_N \leqslant p_0$ and we count the particles (i.e., we measure the number of particles in this configuration), then all the probabilities $p_N(k)$ of all possible outcomes k are smaller than the threshold p_0, and thus, no outcome is possible.

So, only N-particle states for which $1/\sigma_N = 1/(\sigma \cdot \sqrt{N}) > p_0$ are physically possible – in the sense that only in such states, we can actually count the overall number of particles. The largest such N is therefore determined by the equation

$$\frac{1}{\sigma \cdot \sqrt{N}} \approx p_0.$$

Since $\sigma \approx p_0^{-1/4}$, we get $N \approx p_0^{-3/2}$. This is the limit up to which we can count.

We expect the world to be cognizable. In particular, it is reasonable to expect that we can, in principle, count the overall number of particles in every object in the world – and, in particular, in the world as a whole. In the physical world, there is a natural limit on how many particles there can be in an object: this number of particles cannot exceed the overall number of particles in the Universe. Thus, both under our "smallest probability" assumption and in modern physics, there are limits of the amounts that we can actually count:

- under the smallest probability assumption, we cannot count more than $\approx p_0^{-3/2}$ different objects;
- in modern physics, we cannot count more objects that there are elementary particles in the Universe.

It is therefore reasonable to equate these two counting limits, i.e., to equate the largest physically possible number of particles with the overall number N of particle in the Universe.

According to the modern astrophysics (see, e.g., [31, 41]), the vast majority of particles in the Universe are the stable particles. Out of these particles, we have $\approx 10^{80}$ baryons (protons and neutrons) – and approximately the same number of electrons. There are also about 10^8 photons per baryon; there are also neutrinos, gravitons, etc. but their numbers are usually assumed to be much smaller. So, the overall number of particles in the Universe is $N \approx 10^{88} = 10^{80}$ baryons $\times 10^8$ photons per baryon. As a result, we get an estimate $p_0 \approx N^{-2/3} \approx 10^{-60}$.

It is worth mentioning that the relation between the overall number of particles N in the Universe and the overall number of time moments $T/\Delta t$ is one of the relations that Dirac noticed in his original papers [8, 9].

Comment. The reader should be warned that, in comparison to the previous two simpler estimates of the smallest probability p_0, this third estimate is even more based on heuristic quantitative ideas and is, therefore, the least convincing of the three.

Besides, the number of particles in the Universe – on which we base this estimate – is known with an even smaller accuracy than the Planck constant or the Universe's lifetime. There may be, for example, more particles related to the hypothetic "dark matter", particles that we do not observe directly because they practically do not interact with normal matter – other than affecting the gravitational fields and hence the cosmology.

4.8. The agreement between three independent estimates for p_0 is promising. Three independent estimates for p_0 lead to the same value $p_0 \approx 10^{-60}$. This may be a good indication that we are on the right track.

4.9. Possible commonsense corollaries. Since we know the value of p_0, we can now make a specific conclusion: events with probability

$p \leqslant p_0 \approx 10^{-60}$ are physically impossible.

Previously, we were talking about physical consequences of this conclusion. Let us now analyze what this conclusion means in common sense terms. If we repeatedly toss a fair coin, or perform any similar random experiment, then we can have several (h) heads in a row. The probability of having h heads in a row is equal to 2^{-h}. Since $10^{-60} \approx 2^{-200}$, the inequality $2^{-h} > p_0 \approx 10^{-60} \approx 2^{-200}$ means that $h > 200$. Thus, we can have less than 200 heads in a row but we cannot have more than 200 heads in a row.

Similarly, if an experiment can have two possible outcomes, we run it three times and get the first outcome in all three, it may be a coincidence: even if in reality the results are random, it is quite possible (probability $2^{-3} = 1/8 \gg p_0$) that we get first outcome in all three experiments. However, if we run the experiment 200 times and we get the same result in all 200 repetitions then this can no longer be a coincidence. In other words, we get – on a physical level – a justification for physical induction. In the following Section 5, we show how to transform this informal argument into a precise theorem.

4.10. Relation to psychology? This result is in good accordance with a known psychological fact, first established by I. P. Pavlov in his famous experiments with dogs, that 200 repetitions of a joint occurrence is sufficient to make a brain recognize the relation (i.e., in Pavlov's terms, acquire a conditional reflex). Indeed, fewer than 200 repetitions may still be an accidental coincidence but more than 200 mean that the the two co-occurring events cannot be independent.

5. Use of Kolmogorov Complexity and Related Notions to Describe Randomness and Abnormality in Real World

5.1. Why the above formalization of the notion of "typical" is not always adequate. In the previous section, we described a seemingly natural formalization of the notion "typical" ("not abnormal"): if a probability of an event is small enough, i.e., $\leqslant p_0$ for some very small p_0, then this event cannot happen.

The problem with this approach is that *every* sequence of heads and tails has exactly the same probability. So, if we choose $p_0 \geqslant 2^{-100}$, we will thus exclude all possible sequences of 100 heads and tails as physically impossible. However, anyone can toss a coin 100 times, and this proves that some such sequences are physically possible.

Historical comment. This problem was first noticed by Kyburg under the name of *Lottery paradox* [28]: in a big (e.g., state-wide) lottery, the probability of winning the Grand Prize is so small that a reasonable person should not expect it. However, some people do win big prizes.

5.2. Kolmogorov's idea: use computational complexity. Crudely speaking, the main problem arises because we select the same threshold p_0 for all events. For example, if we toss a fair coin 100 times then a sequence consisting of all heads should not be possible, and it is a reasonable conclusion because the probability that tossing a fair coin will lead to this sequence is extremely small: 2^{-100}.

On the other hand, whatever specific sequence of heads and tails we get after tossing a coin, this sequence also has the same small probability 2^{-100}. In spite of this, it does not seem to be reasonable to dismiss such sequences.

Several researchers thought about this, one of them A. N. Kolmogorov, the father of the modern probability theory. Kolmogorov came up with the following idea: the probability threshold $t(E)$ below which an event E is dismissed as impossible must depend on the event's complexity. The event E_1 in which we have 100 heads is easy to describe and generate; so for this event, the threshold $t(E_1)$ is higher. If $t(E_1) > 2^{-100}$ then, within this Kolmogorov's approach, we conclude that the event E_1 is impossible. On the other hand, the event E_2 corresponding to the actual sequence of heads and tails is much more complicated; for this event E_2, the threshold $t(E_2)$ should be much lower. If $t(E_2) < 2^{-100}$, we conclude that the event E_2 is possible.

The general fact that out of 2^n equally probable sequences of n 0s and 1s some are "truly random" and some are not truly random was the motivation behind Kolmogorov and Martin-Löf's formalization of randomness (and behind the related notion of Kolmogorov complexity; the history of this discovery is described in detail in [30]).

This notion of Kolmogorov complexity was introduced independently by several people: Kolmogorov in Russia and Solomonoff and Chaitin in the US. Kolmogorov defined complexity $K(x)$ of a binary sequence x as the shortest length of a program which produces this sequence. Thus, a sequence consisting of all 0s or a sequence 010101... both have very small Kolmogorov complexity because these sequences can be generated by simple programs; on the other hand, for a sequence of results of tossing a coin, probably the shortest program is to write print(0101 ...) and thus reproduce the entire sequence. Thus, when $K(x)$ is approximately equal to the length $\mathrm{len}(x)$ of a sequence, this sequence is random, otherwise it is not. (The best source for Kolmogorov complexity is a book [30].)

However, the existing Kolmogorov complexity theory does not yet lead to a formalism describing when low-probability events do not happen; we must therefore extend the original Kolmogorov's idea so that it would cover this case as well.

5.3. Important comment: we may not know probability at all. In the above three physical Examples 1–3, we know the probabilities of different situations. For example, when we toss a coin, we know the exact probabilities of different sequences of heads and tails; in statistical physics, there are known formulas that describe the probability that all the particles accidentally start moving in the same direction, etc. In these situations, "abnormal" events clearly mean low-probability events.

In some cases, however, physicists do not know the probabilities and still talk about "abnormal" situations. In such situations, it is impossible to formalize "abnormal" event as a low-probability event.

A good example of such a situation is cosmology. In this text, we will briefly describe the corresponding situation; for a more detailed description see, e.g., [31]. The simplest possible space-time models are *isotropic* (direction-independent) pseudo-Riemannian spaces, i.e., spaces of the type $\mathbb{R} \times S$, in which the geometry is the same in all directions. In more precise terms, in an isotropic space, for every two spatial points $x \in S$ and $x' \in S$ and for every two *directions* e and e' (unit vectors in the tangent spaces to S at x and x'), there exists an isometry that maps, for every real number t, the point (t, x) into the point (t, x') and the vector e into the vector e'.

In General Relativity Theory, all isotropic solutions of the corresponding partial differential equation (that describe space-time geometry) have a *singularity*: a space-time point where the solution is no longer smooth or even continuous. In physical terms, the singularity point of the standard solutions is what is usually called a Big Bang – the moment of time at which our Universe started, the point at which the radius of the Universe was 0 and the density of matter was therefore infinite.

In the isotropic case, the equations can be simplified to the extent that we have an explicit analytical expression for the solution. For all these isotropic solutions, there is always a singularity. A natural question is: is there a singularity in the real world?

Several non-isotropic analytical solutions to the corresponding equations have been found, some of these solutions have a singularity, some do not. Physicists have shown that for *generic* initial conditions (i.e., for the class of initial conditions that is open and everywhere dense in an appropriate topology), there is a singularity.

From this, physicists conclude that the solution that corresponds to the geometry of the actual world has a singularity (see, e.g., [31]): their explanation is that the initial conditions that lead to a non-singularity solution are abnormal (atypical), and the actual initial conditions must be typical.

This physicists' argument is similar to the arguments they make in a probabilistic case; the difference is that here, we do not know the probability of different initial conditions.

5.4. Towards a new formalization of Kolmogorov's idea. "Abnormal" means something unusual, rarely happening: if something is rare enough, it is not typical ("abnormal"). Let us describe what, e.g., an abnormal height may mean. If a person's height is $\geqslant$ 6 ft, it is still normal (although it may be considered abnormal in some parts of the world). Now, if instead of 6 ft, we consider 6 ft 1 in, 6 ft 2 in, etc., then sooner or later we will end up with a height h_0 such that everyone who is taller than h_0 will be definitely called atypical, abnormal (to be more precise, a person of abnormal height). We may not be sure what exactly value h experts will use as a threshold for "abnormal" but we are sure that such a value exists.

While every person whose height is $> h_0$ is definitely atypical, a person whose height is below h_0 is not necessarily typical: he may be atypical because of some other properties.

For example, we may consider people atypical because of an unusual weight. Similarly, there exists a weight w_0 such that everyone whose weight exceeds w_0 will be called atypical.

Comment. In general, "abnormal" is clearly a *fuzzy*, non-binary notion. A lot of research has gone into formalizing and understanding what we mean by abnormal in our common sense reasoning. In comparison with this vast area of research, the main objective of this section is very narrow: to formalize one specific (binary) aspect of the notion "abnormal" – its use by physicists to indicate events that are physically impossible.

Let us express the above idea is general terms. We have a *universal set*, i.e., the set U of all objects that we will consider. In the above example, U is the set of all people. Some of the elements of the set U are abnormal (in some sense), and some are not. Let us denote the set of all elements that are *typical* (*not abnormal*) by T.

On the set U, we have several decreasing sequences of sets $A_1 \supseteq A_2 \supseteq \ldots \supseteq A_n \supseteq \ldots$ with the property that $\bigcap\limits_n A_n = \emptyset$.

In the height example, A_1 is the set of all people whose height is $\geqslant 6$ ft, A_2 is the set of all people whose height is $\geqslant 6$ ft 1 in, A_3 is the set of all people whose height is $\geqslant 6$ ft 2 in, etc.

In the weight example, A_1 is the set of all people whose weight is $\geqslant 150$ lb, A_2 is the set of all people whose weight is $\geqslant 160$ lb, A_3 is the set of all people whose weight is $\geqslant 170$ lb, etc.

We know that for each of these sequences, if we take a sufficiently large n, then all elements of A_n are abnormal (i.e., none of them belongs to the set T of not abnormal elements). In mathematical terms, this means that for some integer N, we have $A_N \cap T = \emptyset$.

In the case of a coin: U is the set of all infinite sequences $\omega = (\omega_1 \ldots \omega_n \ldots)$ of results of flipping a coin; A_n is the set of all sequences that start with n heads H…H but have some tails T afterwards:

$$A_n = \{\omega \,|\, \omega_1 = \mathrm{H} \;\&\; \ldots \;\&\; \omega_n = \mathrm{H} \;\&\; \exists n_t > n\,(\omega_{n_t} = \mathrm{T})\}.$$

Here, $\bigcap\limits_n A_n = \emptyset$. Therefore, we can conclude that there exists an integer N for which all elements of A_N are abnormal: $A_N \cap T = \emptyset$.

According to mechanics, the result of tossing a coin is uniquely determined by the initial conditions, i.e., by the initial positions and velocities of the atoms that form our muscles, atmosphere, etc. So, if we assume that in our world, only typical (= not abnormal) initial conditions can happen, we can conclude that the actual result ω of tossing a coin again and again is also typical (not abnormal): $\omega \in T$.

Therefore, since for the above N, we have $A_N \cap T = \emptyset$, we conclude that the actual sequence of results of flipping a coin cannot belong to A_N. By definition, the set A_N consists of all the sequences that start with N heads and have at least one tail after that. So, the fact that the actual sequence does not belong to A_N means that if the actual sequence ω starts with N heads, then this sequence ω cannot have any further tails and therefore, will consist of all heads.

In plain words, if we have tossed a coin N times, and the results are N heads, then this coin is biased: it will always fall on heads.

The Cantor set $U = \{\mathrm{H}, \mathrm{T}\}^{\mathbb{N}} = \{0, 1\}^{\mathbb{N}}$ of all binary sequences (used in the coin tossing example) will be one of our main examples of the universal set. Other examples include general metric spaces – such as the space $C([a, b])$ of all continuous functions on $[a, b]$ with a sup norm.

Let us describe the above abnormality idea in mathematical terms [13, 27]. To make formal definitions, we must fix a formal theory $\mathcal{L}$ that has sufficient expressive power and deductive strength to conduct all the arguments and calculations necessary for working physics. For simplicity, in the arguments presented in this paper, we consider ZF, one of the most widely used formalizations of set theory.

Using ZF is a little bit of an overkill; a weaker arithmetic system RCA_0 is believed to be quite sufficient to formalize all of nowadays physics. Our definitions and results will not seriously depend on what exactly theory we choose – in the sense that, in general, these definitions and proofs can be modified to fit other appropriate theories $\mathcal{L}$.

Definition 1. *Let $\mathcal{L}$ be a theory, and let $P(x)$ be a formula from the language of the theory $\mathcal{L}$, with one free variable x for which, in the theory $\mathcal{L}$, there exists a set $\{x \mid P(x)\}$. We will then call the set $\{x \mid P(x)\}$ $\mathcal{L}$-definable.*

Crudely speaking, a set is $\mathcal{L}$-definable if we can explicitly *define* it in $\mathcal{L}$. The set of all real numbers, the set of all solutions of a well-defined equation, every set that we can describe in mathematical terms is $\mathcal{L}$-definable.

This does not mean, however, that *every* set is $\mathcal{L}$-definable: indeed, every $\mathcal{L}$-definable set is uniquely determined by formula $P(x)$, i.e., by a text in the language of set theory. There are only denumerably many words and therefore, there are only denumerably many $\mathcal{L}$-definable sets. Since, e.g., in a standard model of set theory ZF, there are more than denumerably many sets of integers, some of them are thus not $\mathcal{L}$-definable.

A sequence of sets $\{A_n\}$ is, from the mathematical viewpoint, a mapping from the set of natural numbers to set of sets, i.e., a set of all the pairs $\langle n, A_n\rangle$. Thus, we can naturally define the notion of an $\mathcal{L}$-definable sequence:

Definition 2. *Let $\mathcal{L}$ be a theory, and let $P(n,x)$ be a formula from the language of the theory $\mathcal{L}$, with two free variables n (for integers) and x. If, in some model of the theory $\mathcal{L}$, the set $\{\langle n, x\rangle \mid P(n,x)\}$ is a sequence (i.e., for every n, there exists one and only one x for which $P(n,x)$), then this sequence will be called $\mathcal{L}$-definable.*

Our objective is to be able to make mathematical statements about $\mathcal{L}$-definable sets. Therefore, in addition to the theory $\mathcal{L}$, we must have a stronger theory $\mathcal{M}$ in which the class of all $\mathcal{L}$-definable sets is a set – and it is a countable set.

Denotation. *For every formula F from the theory $\mathcal{L}$, we denote its Gödel number by $\lfloor F \rfloor$.*

Comment. A Gödel number of a formula is an integer that uniquely determines this formula. For example, we can define a Gödel number by describing what this formula will look like in a computer. Specifically, we write this formula in LaTeX, interpret every LaTeX symbol as its ASCII code (as computers do), add 1 at the beginning of the resulting sequence of 0s and 1s, and interpret the resulting binary sequence as an integer in binary code.

Definition 3. *We say that a theory $\mathcal{M}$ is stronger than $\mathcal{L}$ if it contains all formulas, all axioms, and all deduction rules from $\mathcal{L}$, and also contains a special predicate* $\mathrm{def}(n, x)$ *such that for every formula $P(x)$ from $\mathcal{L}$ with one free variable, the formula*

$$\forall y\, (\mathrm{def}(\lfloor P(x) \rfloor, y) \leftrightarrow P(y))$$

is provable in $\mathcal{M}$.

The existence of a stronger theory can be easily proven:

Proposition 1. *For $\mathcal{L}$=ZF, there exists a stronger theory $\mathcal{M}$.*

Proof. We will prove that, as an example of such a stronger theory, we can simply take the theory $\mathcal{L}$ plus all countably many equivalence formulas as described in Definition 3 (formulas corresponding to all possible formulas $P(x)$ with one free variable). This theory clearly contains $\mathcal{L}$ and all the desired equivalence formulas, so all we need to prove is that the resulting theory $\mathcal{M}$ is consistent (provided that $\mathcal{L}$ is consistent, of course).

Due to compactness principle, it is sufficient to prove that for an arbitrary finite set of formulas $P_1(x), \ldots, P_m(x)$, the theory $\mathcal{L}$ is consistent with the above reflexion-principle-type formulas corresponding to these properties $P_1(x), \ldots, P_m(x)$.

This auxiliary consistency follows from the fact that for such a finite set, we can take

$$\mathrm{def}(n, y) \leftrightarrow (n = \lfloor P_1(x) \rfloor \,\&\, P_1(y)) \vee \ldots \vee (n = \lfloor P_m(x) \rfloor \,\&\, P_m(y)).$$

This formula is definable in $\mathcal{L}$ and satisfies all m equivalence properties. The proposition is proven.

Important comments. 1) In the following text, we will assume that a theory $\mathcal{M}$ that is stronger than $\mathcal{L}$ has been fixed; proofs will mean proofs in this selected theory $\mathcal{M}$.

2) An important feature of a stronger theory $\mathcal{M}$ is that the notion of an $\mathcal{L}$-definable set can be expressed within the theory $\mathcal{M}$: a set S is $\mathcal{L}$-definable if and only if $\exists n \in \mathbb{N}\, \forall y(\mathrm{def}(n, y) \leftrightarrow y \in S)$.

In the following text, when we talk about definability, we will mean this property expressed in the theory $\mathcal{M}$. So, all the statements involving definability (e.g., the Definition 4 below) become statements from the theory $\mathcal{M}$ itself, *not* statements from metalanguage.

We have already mentioned that a sequence of sets $\{A_n\}$ is, from the mathematical viewpoint, a mapping from the set of natural numbers to set of sets, i.e., a set of all the pairs $\langle n, A_n\rangle$. Thus, the notion of an $\mathcal{L}$-definable sequence of sets can be also described by a formula in the language $\mathcal{M}$. So, the following definition is valid in $\mathcal{M}$:

Definition 4. *Let U be a universal set.*

- *A non-empty set $T \subseteq U$ is called a set of typical (not abnormal) elements if for every $\mathcal{L}$-definable sequence of sets A_n for which $A_n \supseteq A_{n+1}$ for all n and $\bigcap\limits_n A_n = \emptyset$, there exists an integer N for which $A_N \cap T = \emptyset$.*
- *Once a set T of typical elements is fixed, then:*

– *If $u \in T$, we will say that u is typical, or not abnormal.*
– *For every property P, we say that "normally, for all u, $P(u)$" if $P(u)$ is true for all $u \in T$.*

Example. In the above coin example, $U = \{\mathrm{H}, \mathrm{T}\}^{\mathbb{N}}$, and A_n is the set of all the sequences that start with n heads and have at least one tail. The sequence $\{A_n\}$ is decreasing and $\mathcal{L}$-definable, and its intersection is empty. Therefore, for every set T of typical elements of U, there exists an integer N for which $A_N \cap T = \emptyset$. This means that if a sequence $s \in T$ is not abnormal and starts with N heads, it must consist of heads only. In physical terms, it means a random sequence (i.e., a sequence that contains both heads and tails) cannot start with N heads – which is exactly what we wanted to formalize.

Physical comment. To formalize the physicist intuition, we must assume that in addition to the universal set and to the physical equation, we also have a set T of typical elements.

For each universal set U, there are several different sets T with the above property. For example, if the set T has this property, then, as one can check, for every $u \notin T$, the union $T \cup \{u\}$ also has the same property. Therefore, there cannot be a "maximal" set of typical elements. (Other examples that there are several different sets T are given in the Appendix.)

So, a proper mathematical description of a physical theory should consist not only of the corresponding equations but of a pair consisting of these equations and a set T.

5.5. Relation to Kolmogorov complexity. Kolmogorov complexity enables us to define the notion of a *random sequence*, e.g., as a sequence s for which there exists a constant $c > 0$ for which, for every n, the (appropriate version of) Kolmogorov complexity $K(s_{|n})$ of its n-element subsequence $s_{|n}$ exceeds $n - c$. Crudely speaking, c is the amount of information that a random sequence has.

Random sequences in this sense do not satisfy the above definition, and are not in perfect accordance with common sense – because, e.g., a sequence that starts with 10^6 zeros and then ends in a truly random sequence is still random (in the above sense). Intuitively, for "truly random" sequences, c should be small, while for the above counter-example, $c \approx 10^6$. If we restrict ourselves to random sequences with fixed c, we satisfy the Definition 4.

Another natural approach is to formalize Kolmogorov's idea that we described earlier (in Section 5.2): namely, to claim that an event E is impossible if its probability $p(E)$ is smaller than the threshold $t(E)$ depending on the complexity of E's description: $t(E) = f(K(E))$, where $K(E)$ is the complexity (e.g., a version of Kolmogorov complexity) of the description of the event E, and $f(x)$ is an appropriate function.

There are many ways to define Kolmogorov complexity and random sequences [30]; it is therefore desirable to aim for results that are true in as general case as possible. In view of this desire, in the following text, we will not use any specific version of these definitions; instead, we will assume that Definition 4 holds.

The general results that we will prove under this definition can be also applied to different bounded versions of Kolmogorov complexity-related randomness – as long as these versions satisfy our Definition 4.

5.6. The definition of a set of typical elements is consistent: a proof. It is possible to prove that abnormal elements do exist [13]; moreover, we can select T for which abnormal elements are as rare as we want: for every probability distribution p on the set U and for every ε, there exists a set T for which the probability $p(x \notin T)$ of an element to be abnormal is $\leqslant \varepsilon$:

Proposition 2. *In the theory $\mathcal{M}$, the following statement can be proven: If U is a set and if μ is a probability measure on the set U for which all $\mathcal{L}$-definable sets are μ-measurable, then, for every $\varepsilon > 0$, there exists a set T of typical elements that is μ-measurable and for which $\mu(T) > 1 - \varepsilon$.*

Comment. For example, all arithmetic subsets of the interval $[0,1]$ are Lebesgue-measurable, so for an arithmetic theory $\mathcal{L}$ and for the Lebesgue measure μ, every definable set is measurable. It is worth mentioning that some other set theories have non-measurable definable subsets of the set $[0,1]$.

Proof. By definition of $\mathcal{L}$-definability in $\mathcal{M}$, if a sequence of sets $a = \{A_n\}$ is $\mathcal{L}$-definable, then there exists an integer n_0 for which $y \in \{\langle m, A_m\rangle\}_m \leftrightarrow \mathrm{def}(n_0, y)$. Thus, there are at most countably many $\mathcal{L}$-definable decreasing sequences $a = \{A_n\}$ for which $\bigcap\limits_n A_n = \emptyset$. Therefore, we can order all such sequences into a sequence of sequences: $a^{(1)} = \{A_n^{(1)}\}$, $a^{(2)} = \{A_n^{(2)}\}$, ...

For each k, since the sequence $\{A_n^{(k)}\}_n$ is $\mathcal{L}$-definable, every set from this sequence is also $\mathcal{L}$-definable. Thus, for every k and n, the corresponding set $A_n^{(k)}$ is $\mathcal{L}$-definable. In the proposition, we assumed that every $\mathcal{L}$-definable set is μ-measurable. Thus, for every k and n, the set $A_n^{(k)}$ is μ-measurable.

For each of the sequences $a^{(k)}$, since $\bigcap\limits_n A_n^{(k)} = \emptyset$, we have $\mu(A_n^{(k)}) \to 0$ as $n \to \infty$. Hence, there exists an N_k for which $\mu(A_{N_k}^{(k)}) < \varepsilon/2^k$.

Let us show that as T, we can take the complement $U \setminus A$ to the union A of all the sets $A_{N_k}^{(k)}$. Indeed, by our choice of T, for every $\mathcal{L}$-definable decreasing sequence $a^{(k)} = \{A_n^{(k)}\}$, there exists an integer N, namely $N = N_k$, for which $T \cap A_N^{(k)} = \emptyset$.

To complete the proof, we must show that the set T is μ-measurable and $\mu(T) > 1 - \varepsilon$.

Let us first prove that the set T is μ-measurable. Indeed, for each k, the set $A_{N_k}^{(k)}$ is μ-measurable. Therefore, by the properties of measurable sets, the union $A = \bigcup\limits_k A_{N_k}^{(k)}$ is also μ-measurable. Hence, the complement T to this union is also μ-measurable.

Let is now prove that $\mu(T) > 1-\varepsilon$. Indeed, from $\mu(A^{(k)}_{N_k}) < \varepsilon/2^k$, we conclude that $\mu(A) = \mu\left(\bigcup\limits_k A^{(k)}_{N_k}\right) \leqslant \sum\limits_k \mu(A^{(k)}_{N_k}) < \sum\limits_k \varepsilon/2^k = \varepsilon$, and therefore, $\mu(T) = \mu(U \setminus A) = 1-\mu(A) > 1-\varepsilon$. The proposition is proven.

5.7. Ill-posed problems. As the first potential application of the notion of "typical" to physics, we will show that restriction to "typical" ("not abnormal") solutions leads to regularization of ill-posed problems. In order to describe this idea, let us first briefly describe what are ill-posted problems. Readers who are already familiar with this notion can skip this description.

In many applied problems (geophysics, medicine, astronomy, etc.), we cannot directly measure the state s of the system in which we are interested; to determine this state, we therefore measure some related characteristics y, and then use the measurement results $\widetilde{y}$ to reconstruct the desired state s. The problem of reconstructing the state s from the measurement results $\widetilde{y}$ is called the *inverse problem*. Let us give two examples:

- We are often interested in the *dynamics* of a system, e.g., in measuring the value $x(t)$ of the desired physical quantity x in different moments of time. If we cannot measure $x(t)$ directly, we measure some related quantity $y(t)$, and then try to reconstruct the desired values $x(t)$. For example, in case the dependency between $x(t)$ and $y(t)$ is linear, we arrive at a problem of reconstructing $x(t)$ from the equation $y(t) = \int k(t,s)x(s)\,ds + n(t)$, where $k(t,s)$ is a (known) function, and $n(t)$ denote the (unknown) errors of measuring $y(t)$.
- Another example of inverse problems is *image reconstruction* from a noisy image.

Usually, we know how the actual value y of the measured quantities depends on the state s of the system, i.e., we know a mapping $f : S \to Y$ from the set S of all possible states to the set Y of all possible values of y. Since a measurement is never 100% accurate, the actual measurement results $\widetilde{y}$ are (slightly) different from the actual value $y = f(s)$ of the measured quantity y.

Of course, to be able to reconstruct s from y, we must make sure that we are making sufficiently many measurements, so that from $f(s)$, we will be able to reconstruct s uniquely. In mathematical terms, we need the function f to be reversible (1-1). If this function is reversible, then in the ideal case, when the measurements are absolutely accurate (i.e., when

$\widetilde{y} = y$), we will be able to reconstruct the state s uniquely, as $s = f^{-1}(y)$.

Due to the inevitable measurement inaccuracy, the measured value $\widetilde{y}$ is, in general, different from $y = f(s)$. Therefore, if we simply apply the inverse function f^{-1} to the measurement result $\widetilde{y}$, we get $\widetilde{s} = f^{-1}(\widetilde{y}) \neq s = f^{-1}(y)$. If the measurement error is large, i.e., if $\widetilde{y}$ is very distant from y, then, of course, the reconstructed state $\widetilde{s}$ may also be very different from the actual state s. However, it seems natural to expect that as the measurements become more and more accurate, i.e., as $\widetilde{y} \to y$, the reconstructed state $\widetilde{s}$ should also get closer and closer to the actual one: $\widetilde{s} \to s$.

To describe this expectation in precise terms, we need to find the metrics d_S and d_Y on the sets S and Y which characterize the closeness of the states or, correspondingly, of the measurement results; in terms of these metrics, the fact that $\widetilde{y}$ gets "closer and closer to y" can be written as $d_Y(\widetilde{y}, y) \to 0$, and the condition that $\widetilde{s} \to s$ means $d_S(\widetilde{s}, s) \to 0$. For example, to describe how close the two signals $x(t)$ and $x'(t)$ are, we may say that they are ε-close (for some real number $\varepsilon > 0$), if for every moment of time t, the difference between the two signals does not exceed ε, i.e., $|x(t) - x'(t)| \leqslant \varepsilon$. This description can be reformulated as $d_S(x, x') \leqslant \varepsilon$, where $d_S(x, x') = \sup\limits_t |x(t) - x'(t)|$.

In metric terms, we would like $\widetilde{y} \to y$ to imply $f^{-1}(\widetilde{y}) \to f^{-1}(y)$, i.e., in other words, we would like the inverse function f^{-1} to be continuous. Alas, in many applied problems, the inverse mapping f^{-1} is *not* continuous. As a result, arbitrarily small measurement errors can cause arbitrarily large differences between the actual and reconstructed states. Such problems are called *ill-posed* (see, e.g., [39]).

For example, since all the measurement devices are inertial and thus suppress high frequencies, the functions $x(t)$ and $x(t) + \sin(\omega \cdot t)$, where ω is sufficiently big, lead to almost similar measured values $\widetilde{y}(t)$. Thus, one and the same measurement result $\widetilde{y}(t)$ can correspond to two different states: $x(t)$ and $x(t) + \sin(\omega \cdot t)$.

The fact that a problem is ill-posed means the following: if the *only* information about the desired state s comes from the measurements, then we cannot reconstruct the state with any accuracy. Hence, to be able to reconstruct the state accurately, we need to have an *additional* (prior) knowledge about the state.

In some cases, this knowledge consists of knowing which states from the set S are actually possible, and which are not. For example, we may know that not all signals $x(t)$ $(0 \leqslant t \leqslant T)$ are possible but only smooth signals for which the signal itself is bounded by some value M (i.e., $|x(t)| \leqslant M$ for all $t \in [0,T]$) and the rate with which the signal changes is bounded by some bound Δ (i.e., $|\dot{x}(t)| \leqslant \Delta$ for all $t \in [0,T]$). For this type of knowledge, we, in effect, restrict possible states to a proper *subset* $K \subseteq S$ of the original set S. Then, instead of the original function $f : S \to Y$, we only have to consider its restriction $f_{|K} : K \to Y$ to the set. If this restriction has a continuous inverse, then the problem is solved – in the sense that the more accurate the measurements, the closer the reconstructed state to the original one.

It is known that if the set K if *compact*, then for any 1-1 continuous function $g : K \to Y$ its inverse is also continuous. (It is also known that if a set K is not compact, then for some 1-1 continuous function $g : K \to Y$, its inverse is not continuous.) So, one way to guarantee the continuity of the inverse function $f^{-1}_{|K}$ is to require that the set K is compact. For example, the above prior knowledge about the bounds M and Δ characterizes a set K that is compact in the above metric $d_S(x,x') = \sup |x(t) - x'(t)|$.

We will show that if we restrict ourselves to states S that are typical (= not abnormal), then the restriction of f^{-1} will be continuous, and the problem will become well-posed.

Definition 5. *An $\mathcal{L}$-definable metric space (X,d) is called $\mathcal{L}$-definably separable if there exists an everywhere dense sequence $\{x_n\} \subseteq X$ that is $\mathcal{L}$-definable.*

Comment. As an example, we can consider the Euclidean space $\mathbb{R}^n$ in which points with rational coordinates form an $\mathcal{L}$-definable everywhere dense sequence. Other examples are standard spaces from functional analysis, such as the space $C[a,b]$ of all continuous functions $f : [a,b] \to \mathbb{R}$ with the metric $d(f,g) = \sup\limits_{x \in [a,b]} |f(x) - g(x)|$; in this set, we can consider all finite sets of rational-valued pairs (x_i, y_i) for which $a = x_1 < x_2 < \ldots < x_n = b$, and build continuous functions by linear interpolation. The resulting sequence of piecewise-linear function is an $\mathcal{L}$-definable everywhere dense sequence in $C([a,b])$.

Proposition 3. *Let S and Y be $\mathcal{L}$-definably separable $\mathcal{L}$-definable metric spaces, let T be a set of typical elements of S, and let $f : S \to Y$ be a continuous 1-1 function. Then, the inverse mapping $f^{-1} : f(T) \to S$ is continuous for every $y \in f(T)$.*

In other words, if we know that we have observed a typical (not abnormal) state s (i.e., that $y = f(s)$ for some $s \in T$), then the reconstruction problem becomes well-posed. So, if the observations are accurate enough, we get as small guaranteed intervals for the reconstructed state s as we want.

Proof. It is known that if a set K is compact, then for any 1-1 continuous function $K \to Y$, its inverse is also continuous. Thus, to prove our result, we will show that the closure $\overline{T}$ of the set T is compact.

A set K in a metric space S is compact if and only if it is closed, and for every positive real number $\varepsilon > 0$, it has a finite ε-net, i.e., a finite set $K(\varepsilon)$ with the property that every $s \in K$, there exists an element $s(\varepsilon) \in K(\varepsilon)$ that is ε-close to s.

The closure $K = \overline{T}$ is clearly closed, so, to prove that this closure is compact, it is sufficient to prove that it has a finite ε-net for all $\varepsilon > 0$. For that, it is sufficient to prove that for every $\varepsilon > 0$, there exists a finite ε-net for the set T.

If a set T has a ε-net $T(\varepsilon)$, and $\varepsilon' > \varepsilon$, then, as one can easily see, this same set $T(\varepsilon)$ is also a ε'-net for T. Therefore, it is sufficient to show that finite ε-nets for T exist for $\varepsilon = 2^{-k}, k = 0, 1, 2, \ldots$

Let us fix $\varepsilon = 2^{-k}$. Since the set S is $\mathcal{L}$-definably separable, there exists an $\mathcal{L}$-definable sequence $x_1, \ldots, x_i, \ldots$ which is everywhere dense in S. As A_n, we will now take the complement $S \setminus U_n$ to the union U_n of n closed balls $B_\varepsilon(x_1), \ldots, B_\varepsilon(x_n)$ of radius ε with centers in $x_1, \ldots, x_n$. Since the sequences $\{x_i\}$ is $\mathcal{L}$-definable, this description defines the sequence $\{A_n\}$ in $\mathcal{L}$.

Indeed, by definition of $\mathcal{L}$-definability, the fact that a sequence $\{x_i\}$ is $\mathcal{L}$-definable means that there exists a formula $P(n, x)$ for which $\{x_i\} = \{\langle i, x\rangle \mid P(i, x)\}$. Then, $\{A_n\} = \{\langle n, A\rangle \mid Q(n, A)\}$, where $Q(n, A)$ denotes the following formula:

$$\exists\{x_i\} \left((\forall i \in \mathbb{N}\,(P(i, x_i)))\ \&\ A = S \setminus \bigcup_{i=1}^{n} \{y \mid d(y, x_i) \leqslant \varepsilon\} \right).$$

Thus, the sequence $\{A_n\}$ is $\mathcal{L}$-definable.

Clearly, $A_n \supseteq A_{n+1}$. Since x_i is an everywhere dense sequence, for every $s \in S$, there exists an integer n_0 for which $s \in B_\varepsilon(x_{n_0})$ and for which, therefore, $s \in U_{n_0}$ and $x \notin A_{n_0} = S \setminus U_{n_0}$. Hence, the intersection of all the sets A_n is empty.

Therefore, according to the definition of a set of typical elements (Definition 4), there exists an integer N for which $T \cap A_N = \emptyset$. This means that

$T \subseteq U_N$. This, in its turn, means that the elements $x_1, \ldots, x_N$ form an ε-net for T. So, the set T has a finite ε-net for $\varepsilon = 2^{-k}$. The proposition is proven.

Physical comment. To actually use this result, we need an *expert* who will tell us what is abnormal, and whose ideas of what is abnormal satisfy the (natural) conditions described in Definition 4.

5.8. Comment: every physical quantity is bounded.

Proposition 4. *Let U be a universal set, let T be a set of typical elements of U, and let $f : U \to \mathbb{R}$ be an $\mathcal{L}$-definable function. Then, there exists a real number C such that if $u \in U$ is not abnormal, then $|f(u)| \leqslant C$.*

Proof. To prove the proposition, let us take $A_n \stackrel{\text{def}}{=} \{u \,|\, |f(u)| > n\}$. Since the function f is $\mathcal{L}$-definable, the above formula provides an explicit definition of a sequence $\{A_n\}$ is $\mathcal{L}$. Thus, the sequence $\{A_n\}$ is $\mathcal{L}$-definable.

It is easy to check that the sequence A_n decreases and $\bigcap\limits_n A_n = \emptyset$. Hence, by Definition 4, there exists an integer N for which $A_N \cap T = \emptyset$, i.e., for which, once $u \in T$, we have $u \notin A_N$ – i.e., $|f(u)| \leqslant N$. The statement is proven.

Comment. This Proposition is in good accordance with the example with which we started the discussion of abnormality. In that example, we started by arguing that once a person's height (or weight) exceed a certain threshold, then this person should be considered abnormal. This is, in effect, what the Proposition says, for $U =$ the set of all humans and $f(u) =$ the height (correspondingly, weight) of the person u.

In general, measurable physical quantities come from algorithmically described procedures, hence in a reasonable physical theory, these quantities should be definable in terms of the objects. If we now use the physicists' idea that abnormal initial conditions and/or abnormal values of parameters are impossible, then we can conclude that every physical quantity is bounded.

This conclusion is in good accordance with the fact that in physics, many physical quantities which were originally believed to be unbounded are indeed bounded: distances are bounded by the diameter of the Universe, velocities are bounded by the speed of light, masses of elementary particles are bounded [5, 34], etc.

This conclusion is also in line with Kolmogorov's original idea that some natural numbers which are mathematically possible (like $10^{10^{10}}$) are not physically possible and thus, should not be considered as feasible [19].

Of course, this purely existential result can lead to extremely high bounds, much higher than the actual physical bounds but it is nevertheless encouraging that the existence of at least some bounds comes from our main assumption – that the actual physical objects are not abnormal (in the sense of Definition 4).

5.9. One more idea: towards justification of physical induction. Physicists often claim that if sufficiently many experiments confirm a theory, then this theory is correct. The ability to confirm a theory based on finitely many observations is called *physical induction*; see, e.g., [6].

Physical induction is difficult to formalize, because from the purely mathematical viewpoint, the very fact that some event has occurred many times does not mean that in the next moment of time, this event should necessarily occur. In this section, we will show that our assumption – that all the objects are not abnormal – can lead to a justification for physical induction.

In order to provide such a justification, we will start with an informal explanation of what physicists mean by a physical theory, and then show, step by step, how this explanation can be transformed into a formal definition of a physical theory.

5.10. What is a physical theory? Information motivations, formal definitions, and a mathematical result. A physical theory can be described in different terms: in terms of differential equations, in terms of equalities (like energy conservation) or inequalities (like the second law of thermodynamics, according to which the overall entropy cannot decrease).

From the viewpoint of an experimenter, a physical theory can be viewed as a statement about the results of physical experiments. Some of these experiments are consistent with the physical theory, some are not.

For example, a mechanical theory that describes how particles move can be tested by observing the locations of different particles at different moments of time. For such a theory, the result r_i of i-th experiment is the coordinate of the corresponding particle measured with the corresponding accuracy (e.g., 1.2 or 2.35). In more precise terms, the result r_i of i-th experiment is a point from a finite scale of the ruler or some other measuring instrument (or, if the instrument is binary, the sequence of bits that resulted from the corresponding measurement).

Let $\mathcal{R}$ be the set of possible results of all physically used physical instruments. So, we arrive at the following definition:

Definition 6. *Let an $\mathcal{L}$-definable set $\mathcal{R}$ be given. Its elements will be called possible results of experiments.*

Informal motivations (continued). Let us continue with the informal discussion of what is a physical theory. Intuitively, some sequences $r = (r_1, \ldots, r_n, \ldots)$ of measurement results are consistent with the theory, some are not.

For example, special relativity, via its requirement that velocities cannot exceed the speed of light, imposes a condition that the positions r_i and r_{i+1} of the same particle measured at sequential moments of time t_i and t_{i+1} cannot differ by more than $c \cdot |t_{i+1} - t_i|$.

As we have mentioned, it is reasonable to identify a physical theory with the set of all results $\{r_i\}$ of experiments that are consistent with this theory.

So, we arrive at the following definition:

Definition 7. *Let an $\mathcal{L}$-definable set $\mathcal{R}$ of possible results of experiments be given. By $S = \mathcal{R}^{\mathbb{N}}$, we will denote the set of all possible sequences $r_1, r_n, \ldots$, where $r_i \in \mathcal{R}$.*

- *By a physical theory, we mean a subset $\mathcal{P}$ of the set of all infinite sequences S.*
- *If $r \in \mathcal{P}$, we say that a sequence r satisfies the theory $\mathcal{P}$, or, that for this sequence r, the theory $\mathcal{P}$ is correct.*

Informal motivations (continued). In real life, we only have finitely many results $r_1, \ldots, r_n$; so, we can only tell whether the theory is *consistent* with these results or not, i.e., whether there exists an infinite sequence $r_1, r_2, \ldots$ that starts with the given results that satisfies the theory:

Definition 8. *We say that a finite sequence $(r_1, \ldots, r_n)$ is consistent with the theory $\mathcal{P}$ if there exists an infinite sequence $r \in \mathcal{P}$ that starts with $r_1, \ldots, r_n$ and that satisfies the theory. In this case, we will also say that the experiments $r_1, \ldots, r_n$ confirm the theory.*

Informal motivations (continued). It is natural to require that the theory be *physically meaningful* in the following sense: if all experiments confirm the theory, then this theory should be correct.

An example of a theory that is not physically meaningful in this sense is easy to give: assume that a theory describes the results of tossing a coin, and it predicts that at least once, there should be a tail. In other words, this theory consists of all sequences that contain at least one tail. Let us assume that actually, the coin is so biased that we always have heads. Then, the corresponding infinite sequence of the results of tossing this coin consists of all heads and therefore, does not satisfy the given theory.

However, for every n, the sequence of the first n results (i.e., the sequence of n heads) is perfectly consistent with the theory, because $\mathcal{P}$ contains a sequence H...HT..., in which the first n results are H.

Let us describe this idea in formal terms.

Definition 9. *We say that a theory $\mathcal{P}$ is physically meaningful if the following is true for every sequence $r \in S$:*

If for every n, the results of first n experiments from r confirm the theory $\mathcal{P}$, then, the theory $\mathcal{P}$ is correct for r.

Informal motivations (continued). A physical theory is usually described in a "constructive" way. Namely, for a theory to be effective, we must be able to effectively test whether the theory is consistent with the given observations. In other words, we must have a physically implementable algorithm that, given the results of n observations, checks whether these results are consistent with the given theory.

In other words, for every $r_1, \ldots, r_n$, we can effectively check the property that the results $r_1, \ldots, r_n$ are consistent with this theory. This means, in particular, that this property is definable in the corresponding theory $\mathcal{L}$. Thus, it is reasonable to require that the set $\mathcal{P}$ should also be $\mathcal{L}$-definable.

Definition 10. *We say that a theory $\mathcal{P}$ is $\mathcal{L}$-definable if the set $\mathcal{P}$ is $\mathcal{L}$-definable.*

Now, we are ready for the main result of this section. In this case, the universal set consists of all possible infinite sequences of experimental results, i.e., $U = S$. Let $T \subseteq S$ be the set of typical (not abnormal) sequences.

Proposition 5. *Let $\mathcal{R}$ be a set of possible results of experiments, let S be the corresponding set of infinite sequences, let $T \subseteq S$ be a set of typical elements of S, and let $\mathcal{P} \subseteq S$ be a physically meaningful $\mathcal{L}$-definable theory. Then, there exists an integer N such that if a sequence*

$r = \{r_i\} \in T$ is not abnormal and its first N experiments $r_1, \ldots, r_N$ confirm the theory $\mathcal{P}$, then this theory $\mathcal{P}$ is correct on r.

Proof. For every natural number n, let us define A_n as the set of all the sequences $r = (r_1, r_2, \ldots, r_n, \ldots) \in S$ for which the first n experiments $r_1, \ldots, r_n$ confirm $\mathcal{P}$ (in the sense of Definition 8) but $\mathcal{P}$ is not correct for r (in the sense of Definition 7).

Since the theory $\mathcal{P}$ is $\mathcal{L}$-definable, the above description of A_n is a definition within $\mathcal{L}$; thus, the above sequence $\{A_n\}$ is also $\mathcal{L}$-definable.

It is easy to check that $A_n \supseteq A_{n+1}$. Let us show that the intersection of all the sets A_n is empty. We will prove this emptiness by reduction to a contradiction. Let r be a common element of all the sets A_n. By definition of the set A_n, this means that for every n, the first n experiments $r_1, \ldots, r_n$ confirm $\mathcal{P}$, and $\mathcal{P}$ is not correct for r.

We assumed that the theory $\mathcal{P}$ is physically meaningful. By Definition 9 of physical meaningfulness, from the fact that for every n, the first n experiments confirm the theory $\mathcal{P}$, we conclude that the theory $\mathcal{P}$ is correct for r – a contradiction to the fact that $\mathcal{P}$ is not correct for r. This contradiction shows that the intersection of all the sets A_n is indeed empty.

We can now use the definition of a set T of typical elements (Definition 4), and conclude that there exists an integer N for which $A_N \cap T = \emptyset$ – i.e., for which every element $r \in T$ does not belong to A_N. By definition of the set A_N, this means that once for $r = (r_1, \ldots, r_N, \ldots) \in T$, the results $r_1, \ldots, r_N$ of the first N experiments are consistent with the theory $\mathcal{P}$, it is not possible that $\mathcal{P}$ is not correct on r. Thus, $\mathcal{P}$ *is* correct on r. So, if a sequence $r = \{r_i\} \in T$ is not abnormal and its first N experiments $r_1, \ldots, r_N$ confirm the theory $\mathcal{P}$, then this theory $\mathcal{P}$ is correct on r. The proposition is proven.

This result shows that we can *confirm* the theory based on finitely many observations.

Of course, this "finitely many" may be so large a number that from the viewpoint of working physics, this result will be useless. Another reason why this result is not yet physically useful is that the set T is not $\mathcal{L}$-definable and therefore, we do not know a constructive method of finding this constant N.

However, the very fact that, at least on a philosophical level, we have succeeded in making physical induction into a provable theorem, makes us hope that further work in this direction may lead to physically useful results.

Acknowledgments. This work was supported in part by NASA under cooperative agreement NCC5-209, by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant F49620-00-1-0365, by NSF grants EAR-0112968, EAR-0225670, and EIA-0321328, and by Army Research Laboratories grant DATM-05-02-C-0046.

The authors are greatly thankful to the anonymous referees for their valuable suggestions and recommendations. Their meticulous and thorough analysis of our original text and their (literally) line-by-line suggestions on how to improve this text made them both, in effect, this text's co-authors. The authors are also thankful to Edick Hirsch, the editor of this special issue, for his enormous help.

Appendix: Auxiliary Results

First auxiliary result. The first auxiliary result shows non-uniqueness of a set of typical elements.

Proposition 6. *Let U be a unit disk on a plane with a center at a point O, and let T be a set of typical elements of the unit disk U such that $T \neq \emptyset$ and $T \neq \{O\}$. Then, T cannot be rotation-invariant.*

Mathematical comment. The fact that T is not rotation-invariant means that $T \neq g(T)$ for some rotation g. It is easy to prove that if g is $\mathcal{L}$-definable, then $g(T)$ $(\neq T)$ is another set of typical elements.

Physical comment. This results shows that even if a theory itself is invariant relative to some symmetry group, when we add the necessity to describe typical elements, this symmetry may be violated. This conclusion is in good accordance with the physical idea – called *spontaneous symmetry violation* – that even if a theory is symmetric, the solution is sometimes not; see, e.g., [11].

Proof. In our proof, we will use polar coordinates (r, φ) to describe points on a unit disk; here, $0 \leqslant r \leqslant 1$ and $0 \leqslant \varphi \leqslant 2\pi$. For every n, let A_n be the set of all points $\neq O$ for which $|\varphi - \pi| \leqslant 2^{-n}$ and $\varphi \neq \pi$. The sequence $\{A_n\}$ is clearly $\mathcal{L}$-definable – we gave an explicit definition in $\mathcal{L}$. It is easy to check that the sequence $\{A_n\}$ is monotonic, and that its intersection is the empty set. By Definition 4, we therefore conclude that there exists an integer N for which $A_N \cap T = \emptyset$. In particular, this means that T has no common point with the whole radius $\{(r, \varphi) \mid r \neq 0 \,\&\, \varphi = \pi + 2^{-(N+1)}\}$ that is contained in A_N.

If the set T was rotation invariant, every other radius would also have no common points with T; thus, the only point that T can contain is the point O – but we assumed that $T \neq \emptyset$ and $T \neq \{O\}$. The contradiction proves that T cannot be rotation-invariant.

Second auxiliary result. Another result is related to the following idea. Suppose that we have two identical independent particles, and that x is a quantity that characterizes each particle. We will denote the value of this quantity for the first particle by x_1, and for the second particle by x_2. Let X be the set of possible (typical) values of x_1; since the particles are identical, the set of possible values of x_2 is the same set X. Since the particles are independent, it seems reasonable to assume that what values x_1 are possible should not depend on the actual value of x_2, i.e., that all pairs $(x_1, x_2) \in X \times X$ are possible – and the set of typical pairs (x_1, x_2) is thus equal to $X \times X$. We will show that this is not the case:

Proposition 7. *Let $U = [0,1] \times [0,1]$, and let T be an infinite set of typical elements of U. Then, the set T cannot be represented as $X \times X$ for some set $X \subseteq [0,1]$.*

Physical comment. The intuitive meaning of this result can be, e.g., that no two identical particles are completely independent from each other – which is is good accordance with quantum physics [11.]

Proof. Let us consider $A_n = \{(x_1, x_2) \,|\, x_1 \neq x_2 \,\&\, |x_1 - x_2| \leqslant 2^{-n}\}$. This sequence is clearly definable and decreasing and its intersection is an empty set, so, due to Definition 4, there exists an integer N for which $T \cap A_N = \emptyset$.

If $T = X \times X$, this would mean that for every two values $x, x' \in X$, either $x = x'$ or $|x - x'| > 2^{-N}$. In other words, every two different points from the set X would be separated by at least 2^{-N}. On the interval $[0,1]$, there can be at most 2^N such points, so the set X is finite hence $T = X \times X$ is also finite – which contradicts to our assumption that T is an infinite set. The proposition has been proven.

Third auxiliary result. The final auxiliary result deals with the following. In the main text, we only considered a set of typical elements T of a given universal set. It makes sense to also consider typical elements of different subsets of T: e.g., in addition to a typical height of a person, it may be reasonable to consider a typical height of a child or a typical height of a person living in Russia. For that, we must assume that we for some universal set U, we have a mapping that assigns, to every $\mathcal{L}$-definable set

$A \subseteq U$, a non-empty set $T(A)$ that is a set of typical elements for A. A natural question is: what is the relation between sets $T(A)$ assigned to different sets A? For example, if $A \subseteq B$, will we have $T(A) \subseteq T(B)$? Intuitively, the answer is "in general, no": e.g., the set of all pygmies (A) is a subset of all of the set of all humans B; pygmies have non-typical height, so a typical pygmy is not a typical human ($T(A) \not\subseteq T(B)$). In our general definitions, the answer is indeed "no", even in the simplest physical case of real numbers:

Proposition 8. *Let U be the set of all real numbers, and let T be a mapping that maps every $\mathcal{L}$-definable non-empty subset $A \subseteq U$ into a non-empty set $T(A)$ of typical elements w.r.t. A. Then:*

- *there exist non-empty sets A and B for which $A \subseteq B$ but $T(A) \not\subseteq T(B)$;*
- *there exist non-empty sets A and B for which $T(A \cup B) \neq T(A) \cup T(B)$;*
- *there exist non-empty sets A and B for which $T(A \cap B) \neq T(A) \cap T(B)$.*

Proof. Let $A = [0, 1]$. As a sequence A_n, let us take $A_n = (0, 1/n)$. This is a $\mathcal{L}$-definable monotonic sequence with the empty intersection, so, due to Definition 4, there exists an integer N for which $A_N \cap T(A) = \emptyset$. Hence, for $B = A_N$, we have $B \subseteq A$ and $T(B) \subseteq B$, but $T(B) \not\subseteq T(A)$.

For the same A and B, $T(B) \subseteq T(A) \cup T(B)$, but $T(B) \not\subseteq T(A \cup B) = T(A)$, so $T(A \cup B) \neq T(A) \cup T(B)$.

For the same A and B, $T(A) \cap T(B) = \emptyset$ while $T(A \cap B) = T(B) \neq \emptyset$, so $T(A \cap B) \neq T(A) \cap T(B)$. The proposition is proven.

References

1. J. D. Barrow and F. J. Tipler, *The Anthropic Cosmological Principle*. Oxford University Press, Oxford (1986).
2. V. B. Berestetsky and E. M. Lifshits, *Relativistic quantum theory*. Pergamon Press (1974).
3. B. E. Blank, *The Constants of Nature and Just Six Numbers*. — Notices of the American Mathematical Society **51**, No. 4 (2004), 1220–1225.
4. L. Boltzmann, *Bemrkungen über einige Probleme der mechanischen Wärmtheorie*. — Wiener Ber. II, **75** (1877), 62–100.
5. L. Brink and M. Henneaux, *Principles of String Theory*. Plenum Press (1988).
6. C. D. Broad, *Ethics and the history of philosophy*. Routledge and Kegan Paul, London (1952).
7. R. H. Dicke, *Principle of Equivalence and Weak Interactions*. — Rev. Mod. Phys. **29** (1957), 355-ff.

8. P. A. M. Dirac, *The cosmological constants.* — Nature **139** (1937), p. 323.
9. P. A. M. Dirac, *A new basis for cosmology.* — Proceedings of the Royal Society Series A **165** (1938), 199–208.
10. R. P. Feynman, *Statistical Mechanics.* W.A. Benjamin (1972).
11. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures.* Addison-Wesley (1965).
12. R. P. Feynman, *QED: The Strange Theory of Light and Matter.* Princeton University Press, Princeton (1985).
13. A. M. Finkelstein and V. Kreinovich, *Impossibility of hardly possible events: physical consequences.* Abstr. 8th Int'l Congr. Log., Methodology & Philosophy of Science, **5**, pt. 2, Moscow (1987), 23–25.
14. L. Grover, *A fast quantum mechanical algorithm for database search.* — Proc. 28th ACM Symp. on Theory of Computing (1996), 212–219.
15. L. K. Grover, *Quantum mechanics helps in searching for a needle in a haystack.* — Phys. Rev. Lett. **79** (1997), 325–328.
16. L. Grover, *A framework for fast quantum mechanical algorithms.* — Phys. Rev. Lett. **80** (1998), 4329–4332.
17. W. Heisenberg, *Introduction to the unified field theory of elementary particles.* Interscience, London, New York (1966).
18. F. Herrmann and M. Margenstern, *A universal cellular automaton in the hyperbolic plane.* — Theoretical Computer Science **296**, No. 2 (2003), 327–364.
19. A. N. Kolmogorov, *Automata and life*, in: *Cybernetics expected and Cybernetics unexpected*, Nauka publ., Moscow (1968), p. 24 (Russian).
20. M. Koshelev, *Maximum entropy and acausal processes: astrophysical applications and challenges*, in: G. J. Erickson et al. (eds.), *Maximum Entropy and Bayesian Methods*, Kluwer, Dordrecht (1998), 253–262.
21. O. M. Kosheleva and V. Kreinovich, *On the Algorithmic Problems of a Measurement Process.* — Research Reports in Philosophy of Physics, University of Toronto, Ontario, Canada, Department of Philosophy, No. 5 (1978).
22. O. M. Kosheleva and V. Kreinovich, *What can physics give to constructive mathematics*, in: *Mathematical Logic and Mathematical Linguistics*, Kalinin, Russia (1981), 117–128 (in Russian).
23. V. Kreinovich, *On the possibility of using physical processes when solving hard problems.* — Leningrad Center for New Information Technology "Informatika", Technical Report, Leningrad (1989) (in Russian).
24. V. Kreinovich, *Toward Formalizing Non-Monotonic Reasoning in Physics: the Use of Kolmogorov Complexity and Algorithmic Information Theory to Formalize the Notions 'Typically' and 'Normally'*, in: L. Sheremetov and M. Alvarado (eds.), *Proceedings of the Workshops on Intelligent Computing WIC'04 associated with the Mexican International Conference on Artificial Intelligence MICAI'04*, Mexico City, Mexico, April 26–27 (2004), 187–194.
25. V. Kreinovich and I. A. Kunin, *Kolmogorov Complexity and Chaotic Phenomena.* — International Journal of Engineering Science **41**, No. 3–5 (2003), 483–493.
26. V. Kreinovich and I. A. Kunin, *Kolmogorov Complexity: How a Paradigm Motivated by Foundations of Physics Can Be Applied in Robust Control*, in: A. L. Fradkov and A. N. Churilov (eds.), Proc. Int'l Conf. "Physics and Control" PhysCon'2003, St. Petersburg, Russia, August 20–22 (2003), 88–93.

27. V. Kreinovich, L. Longpré, and M. Koshelev, *Kolmogorov complexity, statistical regularization of inverse problems, and Birkhoff's formalization of beauty*, in: A. Mohamad-Djafari (ed.), Bayesian Inference for Inverse Problems, Proc. SPIE, Vol. 3459, San Diego, CA (1998), 159–170.
28. H. E. Kyburg, Jr., *Probability and the logic of rational belief.* — Wesleyan Univ. Press (1961).
29. S. Lem, *Solaris*, Harcourt (2002).
30. M. Li and P. M. B. Vitanyi, *An Introduction to Kolmogorov Complexity.* Springer, N.Y. (1997).
31. C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation*, W. H. Freeman (1973).
32. D. Morgenstein and V. Kreinovich, *Which algorithms are feasible and which are not depends on the geometry of space-time.* — Geombinatorics, **4**, No. 3 (1995), 80–97.
33. M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information.* Cambridge University Press, Cambridge, U.K. (2000).
34. *Particle Data Group.* — Phys. Lett. **170B**, No. 1 (1986).
35. M. Rees, *Just Six Numbers: The Deep Forces that Shape the Universe.* Basic Books (2001).
36. P. Shor, *Algorithms for quantum computation: Discrete logarithms and factoring.* — Proceedings of the 35th Annual IEEE Symposium on Foundations of Computer Science (1994), 124–134.
37. L. G. Taff, *Celestial mechanics: a computational guide for the practitioner.* Wiley (1985).
38. A. Tarski, *A decision method for elementary algebra and geometry.* Second edition, Berkeley and Los Angeles (1951).
39. A. N. Tikhonov and V. Y. Arsenin. *Solutions of ill-posed problems.* V. H. Winston & Sons, Washington, DC (1977).
40. H. M. Wadsworth Jr., *Handbook of statistical methods for engineers and scientists.* McGraw-Hill, N.Y. (1990).
41. Ya. B. Zeldovich and I. D. Novikov, *Relativistic Astrophysics. Part 2. The structure and evolution of the Universe.* The University of Chicago Press, Chicago and London (1983).

Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA

E-mail: vladik@cs.utep.edu

Institute of Applied Astronomy
St. Petersburg, Russia

Поступило 6 декабря 2004 г.