Towards Applying Computational Complexity to Foundations of Physics

Vladik Kreinovich[1] and Andrei M. Finkelstein[2]

[1]Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA
vladik@cs.utep.edu
[2]Institute of Applied Astronomy
St. Petersburg, Russia

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 describe newer similar results that are already helping physicists. We feel that further research may lead to even more interesting physical applications.

1 Introduction

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 one of us (V.K.) were students at the Math Department of St. Petersburg University. V.K. knew some 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 [32], the first order theory of Minkowski (flat) space-time, with the causal relation $(t, x_1, x_2, x_3) \succ (s, y_1, y_2, y_3)$ iff $t - s > \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2}$, 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.

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 me?

Of courses, 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 physics problems – with the hope that some of these problems can be solved by much faster algorithms that now.

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

Why is this interesting? This activity has led to the excitement 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 is 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 seems to make his geometry less desirable, he immediately suggested that this complexity may be a good thing. For example, the fact that the volume of a ball is a complex expression means that we can then compute this complex expression by simply filling a spherical shell with water and measuring the resulting volume.

This positive twist is behind the current interest in quantum computing: the fact that it is very difficult to solve the quantum mechanics equations seems to indicate that the quantum phenomena, when properly arranged, can speed up computations. And indeed, by using quantum computing, we can reduce the time needed to search in an unsorted n-element array from n to $O(\sqrt{n})$ [10, 11, 12], and the time needed to factor large integers from probably exponential to definitely polynomial [30]; see also [28].

A similar positive twist has also been used in complexity related to space-time geometry: Namely, 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 exponentially increases with R: $V(R) \approx K^R$ for some $K > 1$. Thus, in Lobachevsky space-time, we can solve propositional satisfiability in polynomial time: for every n, we can place exponentially many ($\approx K^n$) processors in a ball of radius n, let each of them 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.

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 [18, 27], is described in more detail in [13]. 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 [15, 17, 27].

Why is this interesting? Because not only can physicists can help us, we can help physics as well. When we have a well-established physical theory, like quantum mechanics, then whatever this theory implies, we have to say "Yes, sir!". If the theory implies that fast prediction is impossible, 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 excited Dima and V.K.: 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. Some of the following examples are rather mathematically trivial, others require some proofs – but in all these cases, we gain some new understanding of physical phenomena.

A word of caution is in order: no matter how excited we may feel, 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 proof of concept – and thus, to hopefully encourage further research in this direction.

2 Proof of Concept: Back-of-the-Envelope Calculations

Before dwelling into more serious math, let us show that this idea can work, on the example of Dirac's large numbers. Let us first explain what they are.

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 $c = 1$.)

Fundamental dimensionless constants. Some physical constants are *dimensionless* in the sense that they are independent on 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*.

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 [8]. In particular, such methods are used in quantum field theory, where we add up several first Feynman diagrams [1]; in celestial mechanics [31], 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 [5, 6].

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). 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, $\Delta E \cdot \Delta t \geq \hbar$ (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 $\Delta t \approx \hbar/E = \hbar/(mc^2)$. Thus, 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.

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 $10^{40} \approx 2^{1/\alpha}$. Indeed, as every computer science person knows, $2^{10} \approx 10^3$, hence $10^{40} \approx (10^3)^{13.3} \approx (2^{10})^{13.3} \approx 2^{133}$.

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.

Why? Since Dirac's 1937 paper, physicists have tried to explain this empirical relation; alas, there is still no widely accepted physics-based explanation. Let us show that simple cognizability (= computational complexity) arguments can explain this relation.

Our back-of-the-envelope explanation. According to quantum mechanics, the dynamics of an n-particle quantum system is described by Schrödinger's equation $i\hbar\dfrac{\partial \Psi}{\partial t} = H\Psi$, where $\Psi(x_1, \ldots, x_n)$ is a *wave function*, a function whose values depend on the coordinates $x_1, \ldots, x_n$ of all n particles.

If some particles do not interact with each other, then we can consider simplified wave functions: e.g., if particles $1, \ldots, m$ and particles $m+1, \ldots, n$ form two clusters that do not interact with each other, then $\Psi(x_1, \ldots, x_n) =$

$\Psi_1(x_1, \ldots, x_m) \cdot \Psi_2(x_{m+1}, \ldots, x_n)$. However, if all n particles actively interact with each other, we need a general function of n coordinate variables.

If we assume that the world is cognizable, then 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? We want to be able to predict something about each of these particles. Thus, for each of n particles i, we must consider at least 2 different locations $x_i^{(1)}$ and $x_i^{(2)}$. In the minimal case, in which we consider exactly two locations for each particle, we still have 2^n possible combinations $(x_1^{(\varepsilon_1)}, \ldots, x_n^{(\varepsilon_n)})$, where $\varepsilon_i \in \{1, 2\}$. Thus, we must consider at least 2^n different values of $\Psi(x_1, \ldots, x_n)$.

Our prediction algorithm must handle each value at least once, so it requires at least 2^n computational steps. The shortest computational step time 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.

3 A More Detailed Analysis (Still on the Physical Level of Rigor)

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 following 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.

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. How-

ever, the physicists say plainly that such an abnormal event is *impossible* (see, e.g., [7]).

Example 2. Another example from statistical physics: Suppose that we have a two-chamber camera. The left chamber if 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, but physicists believe this abnormal event to be impossible. This is a general example of what physicists call *irreversible processes*: 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 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 19 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." [2].

Example 3. If we flip 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, while an engineer (and 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, the physicists (implicitly or explicitly) require that the actual values of the fields must not satisfy the equations, but they must also satisfy the additional condition: that the initial conditions should *not* be *abnormal*.

A natural formalization of this idea. This property of being "not abnormal" has a natural formalization: if a probability of an event is small enough, i.e., $\le p_0$ for some very small p_0, then this event cannot happen. For example, the probability that a fair coin falls heads 100 times in a row is 2^{-100}, so, if we choose $p_0 \ge 2^{-100}$, 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.

What is 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 $\Delta x = c \cdot \Delta \approx \hbar/(mc)$.

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 pointwise 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 $|\Psi(x)|^2$. Thus, the probability p to find a particle in a zone of linear size ε around a point x is $\approx |\Psi(x)|^2 \cdot \varepsilon^3$.

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 threshold p_0, so we will not see anything at all [16].

Comment. This is similar to what Chris Kelvin, the hero of Lem's "Solaris" [24], 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.

We would like to 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 $\rho(x) = |\Psi(x)|^2$ is isotropic Gaussian

$$\rho(x) = \frac{1}{(\sqrt{2\pi})^3 \cdot \sigma^3} \cdot \exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right),$$

with the standard deviation σ equal to the particle's spatial range $\Delta x = \hbar/(mc)$. The largest possible value of the probability density $\rho(x) = |\Psi(x)|^2$ is attained at the central point $x = x_0$ and is equal to $\approx \sigma^{-3} = (\Delta x)^{-3}$. Thus, when $(\Delta x)^{-3} \cdot \varepsilon^3 < p_0$, we conclude that $p < p_0$ for all points x – hence, within this spatial size, we will not be able to see anything at all.

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 $\varepsilon_0 = p_0^{1/3} \cdot \Delta x$.

What is the smallest physically possible probability p_0? How can we find the value p_0? 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., [26]): the so-called Planck length $\varepsilon_0 = \sqrt{\hbar \cdot \gamma / c^3} \approx 10^{-33}$ cm, 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 $p_0 \approx 10^{-60}$.

Lifetime of a free particle – and another way of estimating p_0. According to Schrödinger's equation (see, e.g., [8]), 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 $\rho(x) \cdot \Delta x^3$; similar to localization case, we can conclude that the largest possible value of $\rho(x)$ is $\approx (\Delta x(t))^{-3}$. Thus, 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. Thus, we can determine the lifetime T of a free particle as the largest value T for which the corresponding probability is still $\geq p_0$, 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 $cT/\Delta x = T/\Delta t = p_0^{-2/3}$.

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 $p_0 \approx 10^{-60}$ – exactly the same value as before.

Number of particles in the Universe, one more Dirac's relation, and the third estimate of p_0. In the previous analysis, we considered individual particles. In quantum physics, specifically in field theory, it is known that the real state of the world may not have a fixed number of particle, we may have different number of particles with different probabilities.

We have assumed that events with a low probability ($< p_0$) cannot occur. From this viewpoint, 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, 2, 3, etc., is $\leq p_0$. In a typical such state, the probability $p(k)$ of having k particles is equal to $\approx p_0$ for $k = 0, 2, 3, \ldots$ until some value n_0. This value n_0 can be determined by the fact that in this state, the expected number of particles $\bar{n} = \sum\limits_{k=0}^{n_0} k \cdot p(k) \approx 1$. Since $p(k) = p_0$, this expected value is equal to $1 \approx p_0 \cdot \sum\limits_{k=0}^{n_0} k \approx p_0 \cdot n_0^2/2$, hence $n_0 \approx p_0^{-1/2}$.

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

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

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, when $1/\sigma_N < p_0$ and we measure the number of particles in this configuration, the probabilities $p_N(k)$ of all possible outcomes k are smaller than the threshold, and thus, no outcome is possible.

So, only N-particle states for which $1/\sigma_N = 1/(\sigma \cdot \sqrt{N}) \geq p_0$ are physically possible. The largest such N is therefore determined by the equation

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

Since $\sigma \approx p_0^{-1/4}$, we get $N \approx p_0^{-3/2}$.

It is reasonable to equate the largest physically possible number of particles with the overall number of particle in the Universe $N \approx 10^{88} = 10^{80}$ 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 [5, 6].

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 is a very good indication that we are on the right track.

Possible commonsense corollaries. Since we know the value of p_0, we can now make a specific conclusion: events with probability $p < p_0 \approx 10^{-60}$ are physically impossible.

Previously, we were talking about the 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 $10^3 \approx 2^{10}$, the inequality $2^{-h} < p_0 = 10^{-60}$ means that $2^{-h} < (10^3)^{-20} \approx (2^{10})^{-20} = 2^{-200}$, i.e., it 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 3, 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 next section, we show how to transform this informal argument into a precise theorem.

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.

4 Ideas Related to Computational Complexity Lead to a Consistent Description of the Above Physical Idea and to Useful Physical Applications

Why the above formalization of the notion of "not abnormal" is not always adequate. In the previous section, we described a seemingly natural formalization of the notion "not abnormal": if a probability of an event is small enough, i.e., $\le p_0$ 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 $p_0 \ge 2^{-100}$, we will thus exclude all possible sequences of heads and tails as physically impossible. However, anyone can toss a coin 100 times, and this prove that some sequences are physically possible.

Historical comment. This problem was first noticed by Kyburg under the name of *Lottery paradox* [23]: in a big (e.g., state-wide) lottery, the probability of winning the Grand Prize is so small, then a reasonable person should not expect it. However, some people do win big prizes.

Kolmogorov's idea: use computational complexity. Crudely speaking, the main problem is with selecting 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 follow-

ing idea: the threshold below which the event is dismissed as impossible must depend on the event's complexity. The event in which we have 100 heads is easy to describe and generates, so for this event, the threshold is higher, and the probability 2^{-100} means that we this event is impossible. On the other hand, the event corresponding to an actual sequence of heads and tails is much more complicated; for this event, the threshold is much lower, so the event with a probability 2^{-100} is still possible.

This idea – and the related notion of Kolmogorov complexity – was used by Kolmogorov and Martin-Löf in their formalization of randomness. This notion of Kolmogorov complexity was introduced independently by several people: Kolmogorov in Russia and Solomonoff and Chaitin in the US. Kolmogorov defined a 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 0's or a sequence 010101... have a very small Kolmogorov complexity because these sequences can be generated by simple programs, while for a sequence of results of tossing a coin, probably the shortest program is to write print(0101...) and then 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 [25].)

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.

Important comment: we may not know probability at all. In the above three physical examples when physicists talks about "not abnormal" initial conditions, we knew something about probability. However, there are examples of this type of reasoning in which probability does not enter into picture at all.

For example, in general relativity, it is known that for almost all initial conditions (in some reasonable sense) the solution has a singularity point. From this, physicists conclude that the solution that corresponds to the geometry of the actual world has a singularity (see, e.g., [26]): the reason is that the initial conditions that lead to a non-singularity solution are abnormal (atypical), and the actual initial conditions must be not abnormal.

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 ≥ 6 ft, it is still normal (although it may be considered abnormal in some parts of the world). Now, if instead of 6 pt, we consider 6 ft 1 in, 6 ft 2 in, etc, then sooner or later we will end up with a height h such that everyone who is higher than h will be definitely called a person of abnormal height. We may not be sure what exactly value h experts will call "abnormal", but we are sure that such a value exists.

Let us express this idea is general terms. We have a *Universe of discourse*,

i.e., a set U of all objects that we will consider. 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 a decreasing sequence of sets $A_1 \supseteq A_2 \supseteq \ldots \supseteq A_n \supseteq \ldots$ with the property that $\cap A_n = \emptyset$. In the above example, U is the set of all people, A_1 is the set of all people whose height is ≥ 6 ft, A_2 is the set of all people whose height is ≥ 6 ft 1 in, A_2 is the set of all people whose height is ≥ 6 ft 2 in, etc. We know that 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 n, we have $A_n \cap T = \emptyset$.

In case of a coin: U is the set of all infinite sequences of results of flipping a coin; A_n is the set of all sequences that start with n heads but have some tail afterwards. Here, $\cap A_n = \emptyset$. Therefore, we can conclude that there exists an n for which all elements of A_n are abnormal. According to mechanics, the result of flipping a coin is uniquely determined by the initial conditions, i.e., on the initial positions and velocities of the atoms that form our muscles, atmosphere, etc. So, if we assume that in our world, only not abnormal initial conditions can happen, we can conclude that for some n, the actual sequence of results of flipping a coin cannot belong to A_n. The set A_n consists of all elements that start with n heads and a tail after that. So, the fact that the actual sequence does not belong to A_n means that if an actual sequence has n heads, then it will consist of all heads. In plain words, if we have flipped a coin n times, and the results are n heads, then this coin is biased: it will always fall on heads.

Let us describe this idea in mathematical terms [9, 22]. To make formal definitions, we must fix a formal theory: e.g., the set theory ZF (the definitions and results will not depend on what exactly theory we choose). A set S is called *definable* if there exists a formula $P(x)$ with one (free) variable x such that $P(x)$ if and only if $x \in S$.

Crudely speaking, a set is definable if we can *define* it in ZF. The set of all real numbers, the set of all solutions of a well-defined equations, every set that we can describe in mathematical terms is definable.

This does not means, however, that *every* set is definable: indeed, every 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 definable sets. Since, e.g., there are more than denumerably many set of integers, some of them are thus not definable.

Definition 1. *A sequence of sets $A_1, \ldots, A_n, \ldots$ is called* definable *if there exists a formula $P(n, x)$ such that $x \in A_n$ if and only if $P(n, x)$.*

Definition 2. *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 definable sequence of sets A_n for which $A_n \supseteq A_{n+1}$ and $\cap A_n = \emptyset$, there exists an N for which $A_N \cap T = \emptyset$.*

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

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. 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 above definition.

There are many ways to define Kolmogorov complexity and random sequences [25]; 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 2 is true.

It is possible to prove that abnormal elements do exist [9]; 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 $\leq \varepsilon$:

Proposition 1. *For every probability measure μ on a set U (in which all definable sets are measurable), and for every $\varepsilon > 0$, there exists a set T of typical elements for which $\mu(T) > 1 - \varepsilon$.*

Proof. Similarly to the above argument, one can show that there are no more than countably many definable sequences of sets $\{A_n\}$. Thus, there are at most countably many definable decreasing sequences $a = \{A_n\}$ for which $\cap A_n = \emptyset$. Therefore, we can order all such sequences into a sequence of sequences: $a^{(1)} = \{A_n^{(1)}\}$, $a^{(2)} = \{A_n^{(2)}\}$, ... For each of these sequences $a^{(k)}$, since $\cap A_n^{(k)} = \emptyset$, 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 definable decreasing sequence $a^{(k)} = \{A_n^{(k)}\}$, there exists an N, namely $N = N_k$, for which $T \cap A_N^{(k)} = \emptyset$.

To complete the proof, we must show that $\mu(T) > 1 - \varepsilon$. Indeed, from $\mu(A_{N_k}^{(k)}) < \varepsilon/2^k$, we conclude that $\mu(A) = \mu(\cup A_{N_k}^{(k)}) \leq \sum \mu(A_{N_k}^{(k)}) < \sum \varepsilon/2^k = \varepsilon$, and therefore, $\mu(T) = \mu(U \setminus A) = 1 - \mu(A) > 1 - \varepsilon$.

First application to physics: restriction to "not abnormal" solutions leads to regularization of ill-posed problems. An ill-posed problem arises when we want to reconstruct the state s from the measurement results r. Usually, all physical dependencies are continuous, so, small changes of the state s result in small changes in r. In other words, a mapping $f : S \to R$ from the set of all states to the set of all observations is continuous (in some natural topology). We consider the case when the measurement results are (in principle) sufficient to reconstruct s, i.e., the case when the mapping f is 1-1. That the problem is ill-posed means that small changes in r can lead to huge changes in s, i.e., that the inverse mapping $f^{-1} : R \to S$ is *not* continuous.

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

Definition 3. *A definable metric space (X, d) is called* definably separable *if there exists a definable everywhere dense sequence $x_n \in X$.*

Proposition 2. *Let S be a definably separable definable metric space, T be a set of all not abnormal elements of S, and $f : S \to R$ be a continuous 1-1 function. Then, the inverse mapping $f^{-1} : R \to S$ is continuous for every $r \in f(T)$.*

In other words, if we know that we have observed a not abnormal state (i.e., that $r = f(s)$ 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 $K \to R$, 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 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-set 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 R.

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 definably separable, there exists a definable sequence $x_1, \ldots, x_i, \ldots$ which is everywhere dense in S. As A_n, we will now take the complement to the union U_n of n closed balls $B_\varepsilon(x_1), \ldots, B_\varepsilon(x_n)$ of radius ε with centers in $x_1, \ldots, x_n$.

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

Therefore, according to the definition of a set of typical elements, there exists an 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 $\varepsilon = 2^{-k}$.

Comment. To actually use this result, we need an *expert* who will tell us what is abnormal, and whose ideas of what is abnormal satisfies the (natural) conditions described in Definition 2.

Another application: every physical quantity is bounded.

Proposition 3. *If U is a definable set, and $f : U \to R$ is a definable function, then there exists a number C such that if $u \in U$ is not abnormal, then $|f(u)| \le C$.*

Proof. We can take $A_n \stackrel{\text{def}}{=} \{u \,|\, |f(u)| > n\}$; then, $\cap A_n = \emptyset$, hence there exists N for which $A_N \cap T = \emptyset$, i.e., for which, once $u \in T$, we have $u \notin A_N$ – i.e., $|f(u)| \le N$. The statement is proven.

Measurable physical quantities come from an 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 make the following conclusions:

Special relativity. If as U, we take the set of all the particles, and as f, we take velocity, then we can conclude that the velocities of all (not abnormal) particles is bounded by some constant C. This is exactly what special relativity says, with the speed of light as C.

Cosmology. If we take the same state U, and as f, take the distance from the a particle u to some fixed point in the Universe, then we can conclude that the distances between particles in the Universe are bounded by a constant C. So, the Universe is *finite*. Similarly, if we take a time interval between the events as f, we can conclude that the Universe has a *finite lifetime*.

Why particles with large masses do not exist. Several existing particle classification schemes allow particles with arbitrarily large masses [3]. E.g., in Regge trajectory scheme, particles form families with masses $m_n = m_0 + n \cdot d$ for some constants m_0 and d: when $n \to \infty$, we have $m_n \to \infty$. However, only particles with relatively small masses have been experimentally observed (see, e.g., [29]).

These particles with large masses, that are difficult to weed out using equations only, can be easily weeded out if use the notion of "not abnormal". Indeed, if we take mass of the particle as f, then we can conclude that the masses of all (not abnormal) particles are bounded by some constant C.

Dimensionless constants are usually small. As we have mentioned, this 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.

Thus, complexity idea not only helps us formalize the way physicists think, it also helps us formalize Dirac's intuition that why large constants are abnormal.

This development is in line with Kolmogorov's original idea that some natural numbers which are mathematically possible (like $10^{10^{10}}$) are not feasible and thus, should not be considered as feasible [14].

One more application: justification of physical induction. From the viewpoint of an experimenter, a physical theory can be viewed as a statement about the results of physical experiments. If we had an infinite sequence of experimental results $r_1, \ldots, r_n, \ldots$, then we will be able to tell whether the theory is correct or not. So, a theory can be defined as a set of sequences $r_1, r_2, \ldots$ that are consistent with its equations, inequalities, etc. 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 is an infinite sequence $r_1, r_2, \ldots$ that starts with the given results that satisfies the theory.

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 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, this infinite sequence 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 we can add a tail to it and get an infinite sequence that belongs to the set $\mathcal{T}$. Let us describe this idea in formal terms.

Definition 4. *Let a definable set R be given. Its elements will be called* possible results of experiments. *By S, we will denote the set of all possible sequences $r_1, r_n, \ldots$, where $r_i \in R$. By a* theory, *we mean a definable subset $\mathcal{T}$ of the set of all infinite sequences S. If $r \in \mathcal{T}$, we say that a sequence r* satisfies *the theory $\mathcal{T}$, or, that for this sequence r,* the theory $\mathcal{T}$ is correct.

Comment. A theory is usually described by its axioms and deduction rules. The theory itself consists of all the statements that can be deduced from the

axioms by using deduction rules. In most usual definitions, the resulting set is r.e. – hence definable. We therefore define a theory as a definable set.

Definition 5. *We say that a finite sequence $(r_1, \ldots, r_n)$ is* consistent *with the theory $\mathcal{T}$ if there exists an infinite sequence $r \in \mathcal{T}$ that starts with $r_1, \ldots, r_n$ and that satisfies the theory. In this case, we will also say that the* first n experiments confirm the theory.

Definition 6. *We say that a theory $\mathcal{T}$ 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 $\mathcal{T}$, then, the theory $\mathcal{T}$ is correct for r.*

In this case, the universal set consists of all possible infinite sequence of experimental results, i.e., $U = S$. Let $T \subseteq S$ be the set of all typical (not abnormal) sequences.

Proposition 4. *For every physically meaningful theory $\mathcal{T}$, there exists an integer N such that if a sequence $r \in S$ is not abnormal and the first N experiment confirm the theory $\mathcal{T}$, then this theory $\mathcal{T}$ is correct.*

Idea of the proof: as A_n, we take the set of all the sequences r for which either the first n experiments confirm $\mathcal{T}$ or $\mathcal{T}$ is not correct for r.

This result shows that we can *confirm* the theory based on finitely many observations. The derivation of a general theory from finitely many experiments is called *physical induction* (as opposed to *mathematical induction*). There have been many attempts to justify physical induction. However, in spite of the success, the general physical induction is difficult to justify, to the extent that a prominent philosopher C. D. Broad has called the unsolved problems concerning induction a *scandal of philosophy* [4]. We can say that the complexity-motivated notion of "not abnormal" justifies physical induction by making it a provable theorem (and thus resolves the corresponding scandal). It is also nice to notice that a more "low-brow" thing like computational complexity can be useful in more "high-brow" things like philosophical foundations of physical induction.

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.

References

[1] V. B. Berestetsky and E. M. Lifshits, *Relativistic quantum theory*, Pergamon Press, 1974.

[2] L. Boltzmann, "Bemrkungen über einige Probleme der mechanischen Wärmtheorie", *Wiener Ber. II*, 1877, Vol. 75, pp. 62–100.

[3] L. Brink and M. Henneaux, *Principles of String Theory*, Plenum Press, 1988.

[4] C. D. Broad, *Ethics and the history of philosophy*, Routledge and Kegan Paul, London, 1952.

[5] P. A. M. Dirac, "The cosmological constants", *Nature*, 1937, Vol. 139, p. 323.

[6] P. A. M. Dirac, "A new basis for cosmology", *Proceedings of the Royal Society, Series A*, 1938, Vol. 165, pp. 199–208.

[7] R. P. Feynman, *Statistical Mechanics*, W.A. Benjamin, 1972.

[8] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures*, Addison-Wesley, 1965.

[9] A. M. Finkelstein and V. Kreinovich, "Impossibility of hardly possible events: physical consequences," *Abstr. 8th Int'l Congr. Log., Methodology & Philosophy of Science*, Moscow, 1987, Vol. 5, pt. 2, pp. 23–25.

[10] L. Grover, "A fast quantum mechanical algorithm for database search", *Proc. 28th ACM Symp. on Theory of Computing*, 1996, pp. 212–219.

[11] L. K. Grover, "Quantum mechanics helps in searching for a needle in a haystack", *Phys. Rev. Lett.*, 1997, Vol. 79, pp. 325–328.

[12] L. Grover, "A framework for fast quantum mechanical algorithms", *Phys. Rev. Lett.*, 1998, Vol. 80, pp. 4329–4332.

[13] F. Herrmann and M. Margenstern, "A universal cellular automaton in the hyperbolic plane", *Theoretical Computer Science*, 1003, Vol. 296, No. 2, pp. 327–364.

[14] A. N. Kolmogorov, "Automata and life", In: *Cybernetics expected and Cybernetics unexpected*, Nauka publ., Moscow, 1968, p. 24 (Russian).

[15] 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, pp. 253–262.

[16] 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, 63 pp.

[17] O. M. Kosheleva and V. Kreinovich, "What can physics give to constructive mathematics", In: *Mathematical Logic and Mathematical Linguistics*, Kalinin, Russia, 1981, pp. 117–128 (in Russian).

[18] 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).

[19] 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, pp. 187–194.

[20] V. Kreinovich and I. A. Kunin, "Kolmogorov Complexity and Chaotic Phenomena", *International Journal of Engineering Science*, 2003, Vol. 41, No. 3–5, pp. 483–493.

[21] 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, pp. 88–93.

[22] V. Kreinovich, L. Longprè, 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, pp. 159–170.

[23] H. E. Kyburg, Jr., *Probability and the logic of rational belief*, Wesleyan Univ. Press, 1961.

[24] S. Lem, *Solaris*, Harcourt, 2002.

[25] M. Li and P. M. B. Vitanyi, *An Introduction to Kolmogorov Complexity*, Springer, N.Y., 1997.

[26] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation*, W. H. Freeman, 1973.

[27] D. Morgenstein and V. Kreinovich, "Which algorithms are feasible and which are not depends on the geometry of space-time", *Geombinatorics*, 1995, Vol. 4, No. 3, pp. 80–97.

[28] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information*, Cambridge University Press, Cambridge, U.K., 2000.

[29] Particle Data Group, *Phys. Lett.*, 1986, Vol. 170B, No. 1.

[30] P. Shor, "Algorithms for quantum computation: Discrete logarithms and factoring", *Proceedings of the 35th Annual IEEE Symposium on Foundations of Computer Science*, 1994, pp. 124–134.

[31] L. G. Taff, *Celestial mechanics: a computational guide for the practitioner*, Wiley, 1985.

[32] A. Tarski, *A decision method for elementary algebra and geometry*, 2nd ed., Berkeley and Los Angeles, 1951.