ARTÍCULO

Toward Formalizing Non-Monotonic Reasoning in Physics: the Use of Kolmogorov Complexity

Vladik Kreinovich
Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, TX 79968 (USA)
vladik@utep.edu

Abstract

When a physicist writes down equations, or formulates a theory in any other terms, he or she usually means not only that these equations are true for the real world, but also that the model corresponding to the real world is "typical" among all the solutions of these equations. This type of argument is used when physicists conclude that some property is true by showing that it is true for "almost all" cases. There are formalisms that partially capture this type of reasoning, e.g., techniques based on the Kolmogorov-Martin-Löf definition of a random sequence. The existing formalisms, however, have difficulty formalizing, e.g., the standard physicists' argument that a kettle on a cold stove cannot start boiling by itself, because the probability of this event is too small.

We present a new formalism that can formalize this type of reasoning. This formalism also explains "physical induction" (if some property is true in sufficiently many cases, then it is always true), and many other types of physical reasoning.

Keywords: Non-Monotonic Reasoning, Typical, Kolmogorov Complexity.

1. Introduction

At present, there is sometimes a disconnection between the physicists' intuition and the corresponding mathematical theories.

First, in the current mathematical formalizations of Physics, physically impossible events are sometimes mathematically possible. For example, from the physical and engineering viewpoints, a cold kettle placed on a cold stove will never start boiling by itself. However, from the traditional probabilistic viewpoint, there is a positive probability that it will start boiling, so a mathematician might say that this boiling event is rare but still possible.

Second, in the current formalizations, physically possible indirect measurements are often mathematically impossible. For example, in engineering and in physics, we often cannot directly measure the desired quantity; instead, we measure related properties and then use the measurement results to reconstruct the measured values. In mathematical terms, the corresponding reconstruction problem is called the *inverse problem*. In practice, this problem is efficiently used to reconstruct the signal from noise, to find the faults within a metal plate, etc. However, from the purely mathematical viewpoint, most inverse problems are *ill-*

defined meaning that we cannot really reconstruct the desired values without making some additional assumptions.

A physicist would explain that in both situations, the counter-examples like a kettle boiling on a cold stove or a weird configuration that is mathematically consistent with the measurement results are *abnormal*. In this paper, we show that if we adequately formalize this notion of abnormality, we will be able to weed out these counterexamples and thus, make the formalization of physics better agreeing with common sense and with the physicists' intuition.

The structure of this paper is as follows. In Section 2, we describe the physicists' reasoning that is currently difficult to formalize: namely, their use of notions "typical" and "normal". In Section 3, we briefly explain that this use is different from the standard commonsense uses of these notions and that, therefore, the existing techniques for describing these notions (and non-monotonic reasoning in general) cannot be directly applied to the physicists' use of these notions. In Section 4, we describe a naive commonsense formalization of this use – as impossibility of events with very small probability – and explain that is not fully adequate either.

In Section 5, we present a consistent modification of the (inconsistent) naive approach, a modification based on the ideas of Kolmogorov complexity: namely, that the threshold probability below which an event becomes impossible depends on the complexity of the event's description; this threshold probability should be higher for simple events and much lower for complex events. In the same Section 5, we also explain why this modification is not fully adequate either. In Section 6, we present a new idea and the resulting definitions; auxiliary mathematical details related to these definitions are placed in the Appendix. Sections 7 and 8 describe applications of our new definition: towards the solution of ill-posed problem and towards a more adequate description of physical induction.

2. Physicists' Use of the Notions "Typical" and "Normal"

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., [3]).

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 "microreversible" 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. (It is worth mentioning that irreversibility seems to be related to causality, another topic of interest both to physics and to non-monotonic reasoning.)

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." [1].

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 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 suitable 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 such 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 suitable but if we try to formalize the physicists' opinion more closely, we may sometimes get even better results.

3. Physicists' vs. Common-sense Use of "Typical" and "Normal"

In the previous section, we have seen that to formalize physicists' reasoning, it is necessary to formalize the notions of "abnormal" ("normal") and "typical", especially in a probabilistic context. These notions are traditionally studied in non-monotonic reasoning, since many well-known examples of non-monotonic reasoning are indeed related to these notions, starting with the classical example "Birds normally fly. Tweety is a bird."

There is a massive body of work by J. Pearl, H. Geffner, E. Adams, F. Bacchus, Y. Halpern, and others on probability-based non-monotonic reasoning; many of them are cited in the books [5, 15] that also describes other existing approaches to non-monotonic reasoning, approaches found in the AI and Knowledge Representation communities; see also [6].

The existing approaches have shown that many aspects of non-monotonic reasoning – and, in particular, many aspects of the notions of "abnormal" and "typical" – can indeed be captured by the existing logic-related and probability-related ideas. In this paper, we consider aspects of non-monotonic reasoning that are not captured by the previous formalisms, and we produce a new probability-related formalism for capturing these aspects.

More specifically, our emphasis in this paper is mainly on the notions of "abnormal" and "typical".

4. Naive Description of "Not Abnormal" and Its Limitations

At first glance, it looks like in the probabilistic case, the notion of "not abnormal" has a natural formalization: if the probability of an event is small enough, i.e., $\leq p_0$ 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 $p_0 \geq 2^{-100}$, 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* [12]: 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. Consistent Modification of the Naive Description: Use of Kolmogorov Complexity

5.1. Main Idea

The main problem with the above naive formalization 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 [13]).

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 $\text{len}(x)$ of a sequence, this sequence is random, otherwise it is not. (The best source for Kolmogorov complexity is a book [13].)

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.

There is an alternative (and equivalent) definition of a random sequence which is based on the statistical practice. Namely, in mathematical statistics, we prove, e.g., that with probability 1, the ratio of 1s in a random binary sequence tends to 1/2, and conclude that for a random sequence, this ratio should also be equal to 1/2. In other words, we prove that the set of all the sequences for which the ratio *does not* tend to 1/2 is 0, and from this

fact, conclude that the random sequence does not belong to this set. Thus, it is reasonable to define a sequence to be random if it does not belong to any set of probability measure 0.

Of course, arguments similar to the Lottery paradox show that we cannot use this idea as a literal definition: e.g., the actual "random" sequence ω belongs to a one-element set $\{\omega\}$ of probability 0. However, this problem can be easily resolved: in statistical applications, we are only interested in the sets which describe laws of probability, and these sets must therefore be defined by a formula expressing this law. No matter what mathematical alphabet we use to describe such laws, in every such alphabet, there are only countably many formulas. So, we only need to avoid countably many sets of probability measure 0. It is well known that the union of countably many sets of measure 0 also has measure 0, so by dismissing this union, we get a set of measure 1. Sequence that do not belong to any definable set of measure 0 are called *random in the sense of Kolmorogov-Martin-Löf* (or *KML-random*, for short).

To make the above description a precise definition, we must specify precisely what "definable" means; this will be done later in this text.

Kolmogorov-Martin-Löf's definition of a random sequence is not yet what we need. Indeed, it is know that if we take a sequence α which is random in this sense and add an arbitrary number of zeros in front of this random sequence, then, as one can check, the resulting sequence $0\ldots0\,\alpha$ will also be random.

This property shows that the above notion of a random sequence is not in perfect accordance with common sense. Indeed, e.g., contrary to common sense, 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, $c \approx 10^6$. It is therefore reasonable to restrict ourselves to random sequences with fixed c.

An alternative approach, as we have mentioned, is 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. Then, we can define a sequence ω to be "truly random" if $\omega \notin E$ for all sets E for which $p(E) < f(K(E))$.

5.2. First Limitation: Non-Uniqueness of Kolmogorov Complexity

The above definitions have two serious limitations. The first limitation is related to the fact the above modifications are based on the notion of Kolmogorov complexity, and this notion is not uniquely defined [13]. Let us explain this non-uniqueness in detail.

Kolmogorov complexity of a binary sequence x is defined as the shortest length of a program that generates this sequence. Kolmogorov's definition allows us to use any universal programming language in this definition. It is well known that the length of a program can drastically change when we switch to a different programming language.

This change is not important if we are only interested in the asymptotic properties: e.g., whether there exists a constant c for which $K(s_{|n}) \geq n-c$; see, e.g., [13]. However, the actual value of this constant depends on the language. So, if we use a fixed value c and a language to define which sequences are truly random, this definition will change when we switch to a different language and/or a different constant c.

We do not want a definition of a physically meaning object such as a random sequence to depend on something as artificial (and non-physical) as a choice between Fortran, Java, or C.

5.3. Second Limitation: Need to Go Beyond Probability

In the above three physical examples (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., [14]. 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., [14]): 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.

6. How to Formalize the Notion of "Not Abnormal": A New Approach

"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 $\geq$ 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 in 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 $\geq$ 6 ft, A_2 is the set of all people

whose height is $\geq$ 6 ft 1 in, A_3 is the set of all people whose height is $\geq$ 6 ft 2 in, etc.

In the weight example, A_1 is the set of all people whose weight is $\geq$ 150 lb, A_2 is the set of all people whose weight is $\geq$ 160 lb, A_3 is the set of all people whose weight is $\geq$ 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:

$$A_n =$$

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

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 extremely 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 [4, 7, 8, 11]. 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.

It should be mentioned that in ZF, not for every formula $P(x)$, we have a well-defined set: e.g., due to the well-known Russell's paradox, the set $S \stackrel{\text{def}}{=} \{x \,|\, x \notin x\}$ is not defined. Indeed, if it was well-defined, then $S \in S$ would imply $S \notin S$ and $S \notin S$ would imply $S \in S$ – in both cases, we would have a contradiction.

Using ZF is a little bit of an overkill; a weaker arithmetic system RCA_0 (see, e.g., [16]) is believed to be quite sufficient to formalize all of nowadays physics. In a weaker theory, even fewer sets $\{x \,|\, P(x)\}$ are defined than in ZF.

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 the set $\{x \,|\, P(x)\}$ is defined in the theory $\mathcal{L}$. We will then call the set $\{x \,|\, P(x)\}$* $\mathcal{L}$-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: all these sets are $\mathcal{L}$-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 that 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.

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.

For example, for each version of Kolmogorov complexity and for every constant c, the set T of all binary sequences s for which $K(s_{|n}) \geq n - c$ satisfies Definition 4. This was our motivating example. However, we want our results to be as general as possible. Thus, in the following text, we will not use any specific version of the set T; instead, we will assume that Definition 4 holds.

The general results that we will prove under this definition can be also applied to different *resource-bounded* versions of Kolmogorov complexity-related randomness [13] – as long as these versions satisfy our Definition 4.

To make sure that Definition 4 is consistent, we must prove that abnormal elements do exist; this was proven in [4, 7, 8]. Moreover, we can prove that 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 2. *Let U be a set, and let μ be a probability measure on the set U in 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. In mathematics, a sequence of sets A_n is defined as a function that maps every positive integer n into the corresponding set A_n. A function $f : X \to Y$, in its turn, is defined as a set of pairs $\langle x, f(x)\rangle$. Thus, as we have mentioned earlier, a sequence of sets $\{A_n\}$ is defined as a set of pairs $\{\langle m, A_m\rangle\}_m$ corresponding to different positive integers $m = 1, 2, \ldots$

By definition of definability, a set Y is definable if and only if there exists an integer n_0 for which, for every element y, $y \in Y$ if and only if $\mathrm{def}(n_0, y)$ is true. So, 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: $a^{(1)} = \{A_n^{(1)}\}$, $a^{(2)} = \{A_n^{(2)}\}$, $\ldots$

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

Comment. We started the paper with examples in which physicists talk about "random" elements. Later, we noticed that in other types of physicist reasoning, physicists use a more general notion of "typical" elements. In this section, we provided a definition of a set T of *typical* elements. It is worth mentioning that similar ideas can be used to give the definition of the set R of *random* elements. This definition is given in the Appendix, which also contains the detailed description of the relation between these two definitions and the definition of Kolmogorov-Martin-Löf randomness.

7. Application to 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-posed 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 $\widetilde{s} = f^{-1}(\widetilde{y}) \ne s = f^{-1}(y)$. 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., $|x(t) - x'(t)| \le \varepsilon$. This description can be reformulated as $d_S(x, x') \le \varepsilon$, 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., [17]).

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 $x(t)$ $(0 \le t \le T)$ are possible but only smooth signals for which the signal itself is bounded by some value M (i.e., $|x(t)| \le M$ for all $t \in [0, T]$) and the rate with which the signal changes is bounded by some bound Δ (i.e., $|\dot{x}(t)| \le \Delta$ 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 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 $f^{-1} : Y \to S$ 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 for 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 $-U_n$ 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 sequence $\{x_i\}$ 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 $\{x_i\} = \{\langle i, x\rangle \,|\, P(i, x)\}$. Then, $\{A_n\} = \{\langle n, A\rangle \,|\, Q(n, A)\}$, where $Q(n, A)$ denotes the following formula:

$$\exists\{x_i\}\,((\forall i \in \mathbb{N}\,(P(i, x_i)))\,\&$$

$$A = -\bigcup_{i=1}^{n}\{y \,|\, d(y, x_i) \le \varepsilon\}).$$

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.

8. Application to 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., [2].

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. From this viewpoint, physical induction can be viewed as an example of non-monotonic reasoning: based on the finite number of observations, we conclude that the theory is correct; however, if in the future, a new observation appears that is inconsistent with the theory, we retract this conclusion.

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.

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 described 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.

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.

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.

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.*

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 4. *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.

9. Conclusions and Future Work

When a physicist writes down equations, or formulates a theory in any other terms, he or she usually means not only that these equations are true for the real world, but also that the model corresponding to the real world is "typical" among all the solutions of these equations. This type of argument is used when physicists conclude that some property is true by showing that it is true for "almost all" cases.

There exist formalisms that partially capture this type of reasoning, e.g., techniques based on the Kolmogorov-Martin-Löf definition of a random sequence. The existing formalisms, however, have difficulty formalizing, e.g., the standard physicists' argument that a kettle on a cold stove cannot start boiling by itself, because the probability of this event is too small.

In this paper, we presented a new formalism that can formalize this type of reasoning. This formalism also explains "physical induction" (if some property is true in sufficiently many cases, then it is always true), and many other types of physical reasoning.

In the future, it is desirable combine our new approach with the existing logic-based and probability-based approaches to non-monotonic reasoning.

Acknowledgments

This work was supported in part by the NASA grant NCC5-209, by the AFOSR grant F49620-00-1-0365, by NSF grants EAR-0112968, EAR-0225670,EIA-0321328, and HRD-0734825, by the

ARL grant DATM-05-02-C-0046, by Texas Department of Transportation Research Project No. 0-5453, by the Japan Advanced Institute of Science and Technology (JAIST) International Joint Research Grant 2006-08, and by the Max Planck Institut für Mathematik.

The author is greatly thankful to the anonymous referees for valuable suggestions.

Referencias

[1] L. Boltzmann, Bemrkungen über einige Probleme der mechanischen Wärmtheorie, *Wiener Ber. II*, 75:62–100, 1877.

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

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

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

[5] J.Y. Halpern, *Reasoning about uncertainty*, MIT Press, 2003.

[6] J.Y. Halpern, 'Defaults and Normality in Causal Structures, In: *Proc. of 11th Intl. Conf. on Principles of Knowledge Representation and Reasoning*, Sydney, Australia, September 16–19, 2008 (to appear).

[7] V. Kreinovich, A.M. Finkelstein, Towards Applying Computational Complexity to Foundations of Physics, *Notes of Mathematical Seminars of St. Petersburg Department of Steklov Institute of Mathematics*, 316:63–110, 2004.

[8] V. Kreinovich, A.M. Finkelstein, Towards Applying Computational Complexity to Foundations of Physics, *Journal of Mathematical Sciences*, 134(5):2358–2382 2006.

[9] V. Kreinovich, I.A. Kunin, Kolmogorov Complexity and Chaotic Phenomena, *Int'l J. Engin. Science*, 41(3–5):483–493, 2003.

[10] V. Kreinovich, 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, pages 88–93, 2003.

[11] V. Kreinovich, L. Longprè, 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, pages 159–170, 1998.

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

[13] M. Li, P.M.B. Vitanyi, *An Introduction to Kolmogorov Complexity*, Springer, 1997.

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

[15] J. Pearl, *Causality: Models, Reasoning and Inference*, Cambridge University Press, 2000.

[16] S.G. Simpson. *Subsystems of Second Order Arithmetic, Perspectives in Mathematical Logic,* Springer-Verlag, Berlin, Heidelberg, 1999.

[17] A.N. Tikhonov and V.Y. Arsenin. *Solutions of ill-posed problems,* V. H. Winston & Sons, Washington, DC, 1977.

A. Relation Between "Typical" and "Normal"

We started our paper with examples in which physicists talk about "random" elements. Later, we noticed that in other types of physicist reasoning, physicists use a more general notion of "typical" elements. In the main text, we provided a definition of a set T of *typical* elements. Let us show that similar ideas can be used to give the definition of the set R of *random* elements.

Let us start by recalling the notion of a Kolmorogov-Martin-Löf (KML) random sequence. In the main text, we said that an object is called KML-random if it does not belong to any definable set of measure 0. When we introduced this notion, we did not have a formal definition of a definable set. Now that we have such a definition, we can provide a formal definition of a KML-random sequence:

Definition 11. *Let U be a universal set, and let μ be a probability measure on the set U in which all $\mathcal{L}$-definable sets are measurable. We say that an object $u \in U$ is* KLM-random *if it does not belong to any definable set of μ-measure 0.*

This definition formalizes the notion that if an event has probability 0, then this event cannot happen. In accordance with the discussion in the main text, we would like to formalize a stronger idea: that if an event has a small probability, then it cannot happen. In other words, if we have a sequence of embedded events $A_1 \supseteq A_2 \supseteq \ldots$ whose probability get smaller and smaller (and tends to 0), then one of these events is not possible.

It is known that for the embedded events, the limit probability $\lim \mu(A_n)$ is equal to the probability of the intersection. Thus, the requirement that the probability $\mu(A_n)$ tends to 0 means that the intersection $\bigcap\limits_n A_n$ of these events has probability 0. So, we arrive at the following definition.

Definition 12. *Let U be a universal set, and let μ be a probability measure on the set U in which all $\mathcal{L}$-definable sets are measurable. A non-empty set $R \subseteq U$ is called a* set of random elements *if for every $\mathcal{L}$-definable sequence of sets A_n for which $A_n \supseteq A_{n+1}$ for all n and $\mu\left(\bigcap\limits_n A_n\right) = 0$, there exists an integer N for which $A_N \cap R = \emptyset$.*

Similarly to Definition 4, in which every subset $T' \subset T$ of a set T of typical elements is also a set of typical elements, we can easily conclude that every subset $R' \subset R$ of a set R of random elements is also a set of random elements.

What is the relation between this new definition, the definition of KML-randomness, and Definition 4 of a set of typical elements?

In the main text, we had a pretty general definition of definability. To start answering the above question, we must require that a sequence $A_n = A$ consisting of the same definable element A is also definable. This requirement is clearly satisfied for the standard notion of definability. Under this requirement, the following property holds:

Proposition 5. *Let U be a universal set, let μ be a probability measure on the set U in which all $\mathcal{L}$-definable sets are measurable, and let R be a set of random elements (in the sense of Definition 12). Then, every element $r \in R$ is KLM-random.*

In other words, every random element is KLM-random.

Proof. Let $r \in R$. To prove that r is KLM-random, we must prove that the element r does not belong to any definable set A of measure 0. Indeed, if A is such a set, then the sequence $A_n = A$ is also definable. For this sequence, $A_n \supseteq A_{n+1}$ and $\bigcap\limits_n A_n = A$ hence $\mu\left(\bigcap\limits_n A_n\right) = \mu(A) = 0$. Thus, by Definition 12, there exists an N for which $A_N \cap R = \emptyset$. Since $A_N = A$, we thus conclude that $A \cap R = \emptyset$, hence $r \notin A$. The proposition is proven. ∎

Proposition 6. *Let U be a universal set, and let μ be a $\mathcal{L}$-definable probability measure on the set U in which all $\mathcal{L}$-definable sets are measurable, and let R be a set of random elements (in the sense of Definition 12). Then R is also a set of all typical elements (in the sense of Definition 4).*

In other words, every random element is typical.

Proof. Let R be a set of random elements (in the sense of Definition 12). Let us prove that the set R also satisfies Definition 4. Indeed, let A_n be a $\mathcal{L}$-definable sequence of sets for which $A_n \supseteq A_{n+1}$ for all n and $\bigcap\limits_n A_n = \emptyset$. Then $\mu\left(\bigcap\limits_n A_n\right) = \mu(\emptyset) = 0$ and therefore, by Definition 12, there exists an N for which $A_N \cap R = \emptyset$. The proposition is proven. ∎

The inverse is, in general, not true: a typical element is not necessarily random. For example, let $U = [0, 1]$, and let T be any set of typical elements on U. We have already mentioned that if we add an element to the set T, we still get a set of typical elements. In particular, the set $T' \stackrel{\text{def}}{=} T \cup \{0\}$ satisfies Definition 4. For this set T' of typical elements, the element 0 is typical.

Let us now consider a probability measure that corresponds to a uniform probability distribution of the interval $[0, 1]$ (i.e., to the Lebesgue measure on this interval). The element 0 belongs to a definable set $\{0\}$ of measure 0 and is thus, not KML-random. By Proposition 5, it is thus not random at all.

It turns out that this is an only case when a typical element is not random: every non-KML-random typical element is random.

To formulate and prove this result, we must also require:

- that an intersection of a definable sequence of sets is also definable, and
- that a component-wise difference $A_n - B_n$ between two definable sequences of sets is definable.

These requirements are also satisfied for the natural notions of definability. Under these requirements, the following proposition holds.

Proposition 7. *Let U be a universal set and let T be a set of typical elements (in the sense of Definition 4). Let μ be a probability measure in which all $\mathcal{L}$-definable sets are measurable, and let R_{KML} denote the set of all KML-random element of the set U. Then, the intersection $R \stackrel{\mathrm{def}}{=} T \cap R_{\mathrm{KML}}$ is a set of all random elements (in the sense of Definition 12).*

In other words, an element is random (in the sense of Definition 12) if and only if it is typical and KML-random. So, random elements can be described as typical elements which are also random in the sense of Kolmogorov-Martin-Löf. Thus, to describe all possible sets of random elements, it is sufficient to describe all possible sets of typical elements.

Proof. Let us prove that the intersection r is indeed a set of random elements in the set of Definition 12. Indeed, let A_n be a $\mathcal{L}$-definable sequence of sets A_n for which $A_n \supseteq A_{n+1}$ for all n and $\mu(A) = 0$ for $A \stackrel{\mathrm{def}}{=} \bigcap\limits_n A_n$. Since the sequence A_n is definable, its intersection A is also definable and has measure 0. Thus, by definition of KML-randomness, we have $R_{\mathrm{KML}} \cap A = \emptyset$.

By the requirements on definability, the auxiliary sequence $A'_n \stackrel{\mathrm{def}}{=} A_n - A$ is also definable. It is easy to check that $A'_n \supseteq A'_{n+1}$, and that $\bigcap\limits_n A'_n = \emptyset$. Thus, by Definition 4, there exists an N for which $A'_N \cap T = \emptyset$.

Let us show that for the same index N, we have $A_N \cap R = \emptyset$. Indeed, since the sequence A_n is decreasing, we have $A \subseteq A_N$ and thus, $A_N = A'_N \cup A$. Here, the set A'_N does not have any common elements with the set T – hence with the intersection $R = T \cap R_{\mathrm{KML}}$; the set A also does not have any common elements with the set R_{KML} – hence with the intersection $R = T \cap R_{\mathrm{KML}}$. Thus, the entire union $A_N = A'_N \cup A$ does not have any common points with the intersection R, i.e., indeed, $A_N \cap R = \emptyset$. ■