Probability and Science

R.E. Kalman[*]
Mathematical System Theory
ETH, Zürich, CH=8092 Zuürich, Switzerland

Thank you very much. I'm extremely flattered by this very lengthy and kind introduction. I should begin immediately by saying that my lecture will have almost nothing to do with anything that was said in the introduction, except perhaps the maturity earned through doing the various things that I'm credited with and giving me a certain amount of courage to attack problems which are in some sense much more difficult. But I should emphasize that this is only courage to attack the problems, and not a claim to have succeeded.

In fact, Professor Willems indicated to me (I think fairly strongly) that my lecture should not be overly technical. I was very grateful for that instruction, because I couldn't make my lecture too technical, even if I wanted to. So I will talk in very general terms. I am grateful for this opportunity to be able to do so, because I think the problem in question is an extremely important one. In fact, if successfully attacked, which remains to be seen, its solution will undoubtedly have very profound implications on many, many things.

To briefly review what I will say let me indicate that since, roughly speaking, the end of World War II, there has been a kind of madness sweeping through the less talented part of the scientific world, having to do with translating almost any problem into probabilistic terms. We have probabilistic linguistics, we have probabilistic linear programming, we have probabilistic anything, without asking how we should think of the implications of this work and its ultimate utility, its ultimate importance, etc.

What I will really be concerned with is a question of whether the probabilities could be said to exist in the real world and not only in mathematical textbooks. This question is very basic for, for example, the currently totally unresolved puzzles about quantum mechanics, where we have a probabilistic interpretation but nobody claims that it is the correct one. Even beyond such questions, the very idea that the word probability is properly applied to the real world is not

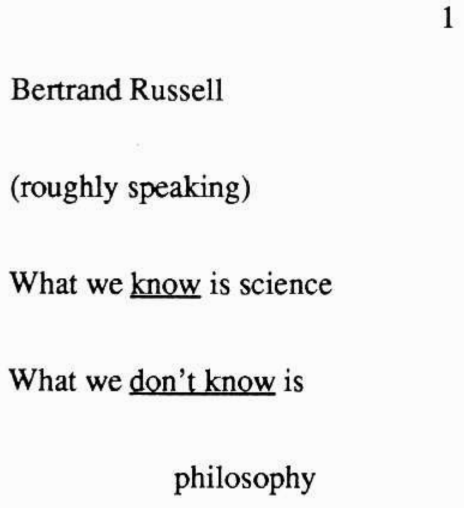
---

[*] Text of the 1991-1992 Johann Bernoulli lecture given by Prof. Kalman at the University of Groningen on June 3, 1992. The Johann Bernoulli Foundation for Mathematics founded in Groningen organizes each year a Johann Bernoulli lecture for which it invites prominent scientists, in particular mathematicians. Johann Bernoulli was professor of mathematics at the University of Groningen from 1695–1705.

at all clear. For example, one applies so-called probabilistic reasoning in the situation like the current U.S. election where we have three candidates. Each candidate gets roughly one third of the vote on a trial basis right now. Therefore people might be tempted to conclude that the probability of either one of the three getting in is approximately one third. This type of reasoning is utter and complete nonsense, has nothing to do with the real meaning of the word probability and it is actually bad for our understanding of some of these subjects.

A much more technical, but also more important area is the question of statistics. It has been said in the past that statistics have been made rigorous by the introduction of probability theory. Unfortunately however, the correct statement is probably the other way around. Statistics have been made irrelevant by the introduction of probability theory. Probability theory is an interesting subject and has considerable intellectual appeal to which some of the Bernoulli's have contributed. But the issue of what it has to do with the real world has in general not been examined.

So this is what I would like to do. To explain perhaps a little better of what I mean by these various terms, let us have the first slide (1):



1

Bertrand Russell

(roughly speaking)

What we know is science

What we don't know is

philosophy

This is a statement that I heard from a friend of mine, who tried to get a paper published in a collection volume, whose editor unfortunately had died in the meantime, otherwise I would name him. The editor was so incensed about Russell's opinion, with which I agree, that he almost refused to publish the paper. Where upon my friend sent me a cassette recording of Russell's own voice, that I could a little bit recognise since I did hear Russell speak while he lived and I can assure you that Russell did mean that statement.

It is simply a means of separating the word "philosophy" from the word "speculation". One of the basic troubles in my topic today is that a lot of the claims made for probability in so-called applications are in the realm of speculation and not in the realm of science. We simply don't know, it might perhaps be right, but in general it is not. It is not unusual to have such worries about so-called "past accomplishments" and in some sense my role model in this lecture is going to be Roger Penrose – let's have the next slide (2):

Roger Penrose (1992)

Gödel's results may not

restrict mathematics if

the process of "proof"

is not mechanical (digital)

but involves creative

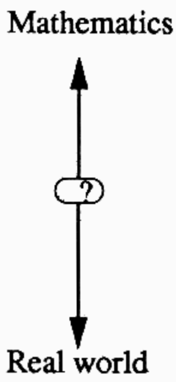
imagination (analog)

[Analog counterrevolution against digital]

– whose lecture I heard last February in Cambridge. He was concerned about a totally different topic, namely the question if it is true that Gödel's results, which are spectacular mathematics, have really something to do with the future of mathematics as a sort of an intellectual pursuit. He concluded that maybe the answer is no. Because in mathematics, real mathematics, living mathematics, mathematics in the real world, not the purely logical structured mathematics, there are examples where people come up with ideas or theorems, proofs of theorems, also theorems themselves, which don't look as though they could be codified in a way that would allow the applicability of the Gödel machinery. You know, if in some sense the method of mathematics is not at all denumerable, then Gödel's arguments are not right, as far as pinpointing the future of mathematics is concerned, and we just have to change our entire thinking.

So my main theme will be that we may have to change our entire thinking about notions surrounding probability. Now of course it is a temptation for me to be able to speak here, literally ex cathedra, on that subject. But I will not do more than raise the issue, propose an obvious remedy and hope that some of you will quickly start working, and proceed to much more and much deeper accomplishments than what I am able to pinpoint at this moment. So let's have the third slide (3):

3



Mathematics

?

Real world

The whole problematique of what I will talk about is really the relationship between mathematics and the real world. Something which has been discussed since the time of the Greeks, but which in my opinion is not adequately understood by any means, even today, even by experts, except people like Penrose. So in the case of probability theory the question is: Probability theory is an interesting mathematical construction – it is possible to make it very precise and very clear and so forth – but, is it true that this sort of a mathematical construction is compatible with the real world in which we live? In particular, whether this is a good way of describing the obvious randomness that we see around us. For example, if we look at a tree, the average tree does not look exactly like your average another tree, even though of the same kind. Therefore we feel that there is some sort of a randomness involved: not all the branches look exactly the same, not all the leaves have exactly the same shape or are the same in number, and so forth. But maybe there is nevertheless some sort of underlying similarity also. So how do we define randomness in the real world?

Now I have only two basic points to make in this connection. The first one is on the next slide (4):

separate

random

from

"probability"

I want to propose that we separate two things, namely randomness and probability. We will therefore, much for the rest of the lecture, primarily talk about randomness and not say very much about exactly how the quantitative description of randomness then leads to the notion of probability. I insist on that because I feel that randomness has not been discussed enough and that probability has been discussed too much.

The second point is that one should sort of try to think of probability in terms of other concepts. Let us look at the next slide (5):

probability
≠ measurable
= system
phenomenon

The first point about probability which we should be worried about, is that it seems that probability is not really measurable. It is not something I can measure like I can measure a resistor obeying Ohm's law. But it is almost always calculated in an indirect way, from some other considerations. For example, if I have a die and a cube, I assume by a certain amount of sloppy thinking or theological conviction that each of the six faces of the cube have the same probability of occuring when I throw the cube and play dice, and therefore I assign probability 1/6 to each one of them. But the important thing is that while this argument appeals to symmetry and so forth, and may be logically perfectly okay, and even that if I could measure the actual probabilities I will detect this kind of thing, it is actually very hard to measure these probabilities in any effective way that imitates the measuring procedures of the physicists.

I should, perhaps before making the second main point, go to the next slide (6):

Dr. Arbuthnot (1713)

$$\text{prob}_{\text{male birth}} :=$$

$$\frac{\#\text{ male births}}{\#\text{ births}} = 0.516$$

which is from perhaps the first published paper on applied probability that is still interesting to read, especially since Daniel Bernoulli made a contribution to that subject about twenty years later.

The problem is this: Somebody observed in London that there are proportionately more boys rather than girls being born, according to whatever records people could come up with. The records go back to about the first third of the 17th century. A member of the Royal Society, Dr. Arbuthnot, had asked an interesting question. The question was: Are the male and female births random? And he calculated the probability - I repeat he *calculated* the probability - of what would happen if you assume that this was a binominal process and compare the probability of the data based on equal probabilities for male and female births. He concluded that the chances of explaining the data with the model of equal probability of male and female is so low that this hypothesis has to be rejected, and therefore he said: "Therefore we know that this process is not random".

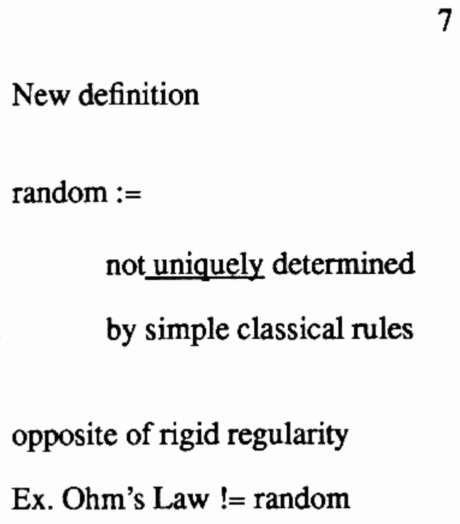
In other words, he equated randomness with perfect randomness. That is to say, since there is no reason to believe that there should be more male than female births, we have to assume that there are equal probabilities, namely 1/2, and that consideration that defines a certain idealized random model simply cannot be reconciled with the data. Indeed, the frequency of occurence was about 0,516 for male births, considerably more than 0,5, and since you have a record over about 80 years, the probability of that many births occuring with equal probability of male and female, was simply much too low, something of the order of 10^{-40}, to be reasonably acceptable.

So we have immediately a big confusion about various things, namely: What do we mean by random? Do we mean by random perfectly random, or something else? Do we mean by probability something that is related to frequency of occurence or not?, etc. But nevertheless, the interesting fact is that in the first published paper on the application of probability theory, one reaches a conclusion that the probabilities that one would like to examine somehow don't exist in nature, because they are incompatible with the data.

Later on, Bernoulli had taken on a more sophisticated point of view. He said: "Why don't we define the probability of male births to be equal to its frequency, namely 0,516, and then recalculate everything". If you do that, of course you get a tremendously good match with the data and you might think you have explained something. However there are still (if you take shorter data runs) big fluctuations in this number, etc., etc.

In other words, it is not a trivial matter, even in such a relatively simple situation, to really measure the probabilities that are hopefully there. In fact, as far as I know, this issue is even today not resolved. Even though of course it can be explained by playing around with the parameters of the model. If you play around with the parameters of any model, then you can explain any data, given enough parameters.

So with such a remark about the difficulty of measuring probability, let me come to my second and most important point, that is on slide number 7:



7

New definition

random :=

not uniquely determined

by simple classical rules

opposite of rigid regularity

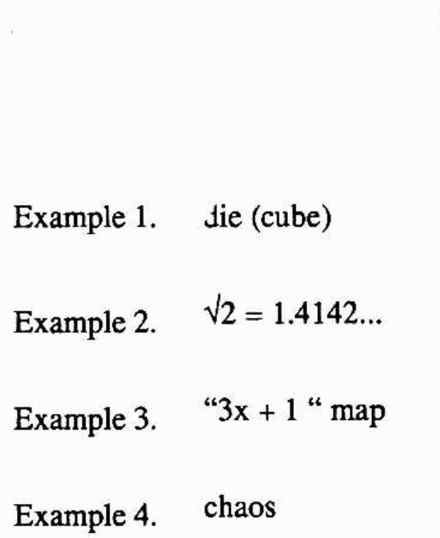
Ex. Ohm's Law != random

After a great deal of thinking about this subject, I came to the conclusion that the only way to define randomness is so that the definition automatically corresponds to reality, but not necessarily to one's intuitive notion of what randomness should be like. Our intuition, of course, is distorted by the 300 years of history of probability theory. On the other hand, if you look at new problems and you look at new data, it is perfectly clear that the only way to define randomness is by saying that randomness is the same thing as a lack of complete repeatability, complete uniqueness in the data.

Now it is an unexplained and very mysterious part of the physical world, that part of the physical world seems to be very unique. For example, Ohm's Law works marvellously and works maybe more than a hundred million times every day in every laboratory, inside every computer and so forth. And there is absolutely nothing random or nonunique about it. However, there are other things which are different. Before trying to understand what the terminology of randomness should mean, we should simply separate the cases. So my definition does that, it separates the completely regular case, which is what mathematicians usually call uniqueness, from another case, which I'm forced to call randomness.

And that is a very nice, technically workable definition. For example, Professor Lions told me about two years ago that in his group of research on partial differential equations, they are generally quite successful in establishing existence theorems. But they have a lot more trouble with uniqueness theorems. If you take a philosophical point of view in the right sense, one might say that the difficulty of proving uniqueness of partial differential equations must be related to the fact that there is something inherently random in the existing solutions of the equations, in the sense that they happen to be nonunique. Somehow the rules of the game that yields partial differential equations are not quite enough to say exactly what should happen and when that is the situation, I call the result random.

Now let me give 4 examples in discussing this definition, which are listed here on slide (8):



8

Example 1. die (cube)

Example 2. $\sqrt{2} = 1.4142...$

Example 3. "$3x + 1$" map

Example 4. chaos

The first example is the simplest: If I take a cube, then by definition, if I throw the cube, the result can have only six possible outcomes, because I define a cube in such a way that it cannot come to rest on a vertex or on an edge. Therefore, by definition it comes to rest on one of the six faces. This applies, of course, to any regular polyhedron. For example, to a Platonian solid. I do not claim from that point of view that the probabilities necessarily will be one over the number of faces. It could be, or it could be proportional perhaps to the area of the faces in less regular situations, and so forth, but exactly how probabilities would come about in the usual sense, while I throw a die, is a much more involved topic to be discussed, because for that I must know that there is not a little magnet inside, etc. But, the fact that the outcome of such an experiment is nonunique, of course not violating any principle of determinism in physics, it is somehow determined by the way I throw the die, the initial conditions and so forth. But nevertheless, practically speaking, the outcome is one of the six possibilities and is therefore not unique. That is what I call random.

I will come back to other, more sophisticated comments as to why this is inevitable. But I think that it illustrates that it is a reasonable definition to make. So I can define random as nonunique without worrying about with what probability the various outcomes will occur.

The second question has to do with the $\sqrt{2}$, which was a mysterious object already to the Greeks and they called it irrational. That is to say, in a certain sense it does not obey what a decent number should look like. Namely, if I compute the $\sqrt{2}$ on a computer, and we have looked at experiments with something like 500 digits a little bit, just to get an idea as to what happens, the intuition says: well, the only thing I can say about each digit in a decimal expansion is that it is some digit between zero and nine, and although I can't predict it from the theoretical set-up so far, there seems to be some evidence that these digits occur with equal probability. But that is not part of my definition. But beyond the fact that all these digits occur, seemingly very little can be said.

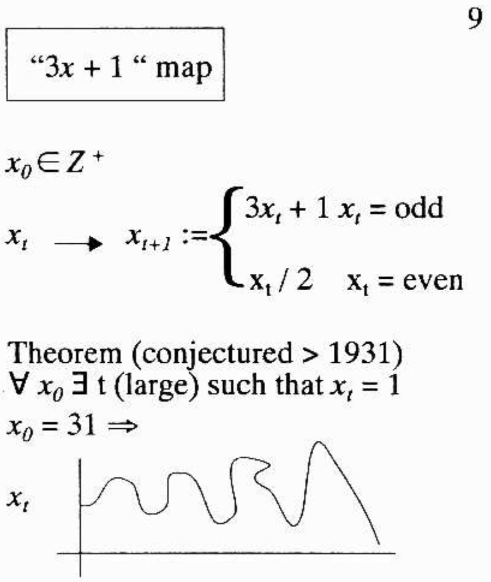
But I would still like to give a proof that this number is random according to my new definition, and the proof is simple. If I look at a sequence of decimal digits, then, according to classical rules, the only thing that could happen in a classical regularity situation, is that the sequence is periodic – a constant sequence of course being a special case – or at least ultimately periodic. If that doesn't happen, then I have an irregular sequence. It is not explained by classical rules, therefore by definition it is random. Now we happen to know that if the sequence of the integers is ultimately periodic, that corresponds to the number that we are expanding being rational number, namely the ratio of two integers. So this is very nice, the ratio of two integers is a perfectly classical, well defined object. Otherwise, the number will happen to be random. So the $\sqrt{2}$ is to be interpreted as random.

There are many problems if you follow this point of view with classical probability theory. First of all, mathematicians may have the ambition of proving, that if I take a very long sequence of these numbers, each digit occurs an equal number of times. I don't know if such a proof has actually been supplied so far, I am sure there has been some attempts for partial results, but I'm not sure whether the issue is settled. But even if it is settled, I don't know very much about the situation from a point of view of probability theory, because probability theory likes to work with things which are independent. So, I would actually like to prove that the occurence of these numbers is independent of whatever went before. But such a proof is ridiculous, because they are obviously not independent, it is a totally determined process and the previous sequence of digits uniquely determines what happens later. So they are certainly not independent in that sense. But I could collect second order statistics, I could look at the correlation of pairs of numbers, of pairs of numbers separated by a certain interval. And it could very well happen that I could prove that these pairs of numbers also behave as though they were independent and so forth. If I engage in such a activity, I waste a great deal of possibly precious brainpower, because the whole objective is meaningless. I cannot prove that something behaves like an independent random process, where on most basic and intuitive grounds, it is not such a process. Therefore, why do it at all? In other words, the assignment of probabilities to the description of the $\sqrt{2}$ is a totally different kind of problem than defining it to be random, according to my new definition, which goes very easily.

Now I should also say that my new definition of randomness is a very optimistic definition, because it is a function of time. Actually finding a simple classical rule, depends on how clever we are. It might happen that somebody discovers a very simple way of describing a complicated looking sequence for the $\sqrt{2}$ and suddenly there is some sort of regularity in that sequence that I simply didn't see before. That is possible. In that case, because of scientific progress, whatever previously was called random, is now not random. And that is perfectly okay. Because the whole problem in the development of science since the time of the Greeks was that certain things seemed to be regular, others not. How can we be sure? What doesn't seem to be too regular, for example the motion of a planet before Copernicus, may be more regular if you take a different point of view. Before Copernicus, if you looked at the planets in an earth centered coordinate system, you would observe a so-called retrogade motion, where a thing goes forward and then comes back in a funny cusp. Well, it is very odd that such a thing should happen, it turns out that it is nothing else but an artifact which is due to a lack of basic knowledge at the time before Copernicus. So therefore, whatever is random according to my definition could very well, with progress, turn out to be nonrandom later on. And that's very nice.

Let's take the next problem, the so-called $3x + 1$ map. I mention that because it is easy enough to remember in spite of the nontechnical details in this lecture and you can immediately try it out on your personal computer at home and perhaps you will resolve the dilemma that is worth at least 2000 dollars in various prizes offered so far, maybe more.

So let's have the next slide (9):



9

$\boxed{\text{“}3x + 1\text{ “ map}}$

$x_0 \in Z^+$

$$x_t \longrightarrow x_{t+1} := \begin{cases} 3x_t + 1 & x_t = \text{odd} \\ x_t / 2 & x_t = \text{even} \end{cases}$$

Theorem (conjectured > 1931)
$\forall x_0 \, \exists \, t$ (large) such that $x_t = 1$
$x_0 = 31 \Rightarrow$

The $3x + 1$ problem is defined as a map, in which I simply iterate a positive integer, not a negative one because then the results are different. I do it in such a way that if the number is odd, I produce from it a an even number, and if the number is even, then I divide it by 2 until it becomes odd. If I try to do this process, for example, if I begin from a number 31 (if I remember correctly), it takes an enormous amount of time before the natural conjecture, namely that eventually we will get the number one, actually occurs. I think that it takes about a 120 iterations in the case of 31. If you look at the partial results of what happens, they look like the graph that I have drawn. This graph certainly has to be called random, because it goes up and down without any obvious reason, although there is a simple rule for it as far as a generating rule is concerned. But remember that my definition of randomness has to do with sort of a global understanding of what goes on. The simple generating rule doesn't count. That still leaves randomness open. So it is a very irregular looking graph. From a point of view of at least my intuition, what we will have to say is that we have a superposition of two phenomena: One is the going to one, and the other is some sort of random going back and forth.

Unfortunately this idea is technically too weak to prove the conjecture that ultimately you always go to one. But it does illustrate the fact that if I don't have enough knowledge about such a thing, then the lack of knowledge or lack of regularity that is there, can best be explained by my definition of randomness. In terms of that definition it is perfectly natural to call this graph random, even though it will ultimately go to one and that is not normally considered random by the usual definition of the word. So, if we had a total understanding of this particular little problem, then we would have a better idea as to how to technically apply the new definition of random. But on an intuitive level, everything is fine.

There is another example that is very embarrassing to conventional probability theory, and that is chaos. Chaos is the discovery that randomness in nature can be grabbed by a certain modelling procedure, which is very different from what one is used to in Euclidean geometry. I am referring to the idea of fractals by Mandelbrot and various consequences of that. But if you look at, for example, fractally generated images of mountains, not only do you have a kind of irregularity that seems to resemble if you just compare pictures of what you see in nature, but it is a kind of irregularity that you can even measure. You can measure a certain fractal exponent and you have a certain control over what is happening, even though this fractal exponent is certainly not the same thing as a probability.

Now why is this an embarrassing example? Well, in probability theory we started out after all from games of chance, it is its only practical interest. Today it might start out from looking at the stock market, which is of practical interest, or a currency exchange market, of an even bigger practical interest. Probability theory started out from such things. If probability had some relevance to the real world, and we are fairly sure that chaos models, although not exactly, the real world, have a lot of relevance to the real world, then there must be an interaction between chaos and probability theory. In practical terms this means that the many professors of probability theory at the universities and the many professors of chaos at the universities will engage in a furious interchange in which people on probability theory write papers saying: "Aha, my probabilistic model explains chaos" and people in chaos write papers, saying: "My chaos models solve problems in classical probability theory" and so forth. There will be a tremendous activity back and forth. Well, if you look at the literature what do you see? You see zero, nothing, no interaction, no papers that have both chaos and probability in the same time. In other words, the classical models of probability coming out of games of chance appear to be irrelevant to the real world, given that one is now fairly convinced that chaos models are relevant to the real world.

Let me add one more comment which I am personally very much interested in, but which is far too technical to explain in the few minutes remaining. And that is the application of these ideas to statistical problems, such as they occur in econometrics. The chief problem in these applications is that the methodology in statistics and econometrics begins with postulating that we know how the real world is. We already have a probabilistic model explaining the irregularity in the data that we look at. We don't ask the data: "Why are you irregular the way you are?", but we assume that we know that and we just try to fit the data. When you do such a thing, you proceed in such a nonscientific way, in other words, you answer your own question before looking at the data. In my opinion, if you proceed this way, there is almost no chance for success, except in a very few cases where you are lucky enough to guess right. It is possible to guess right. But it is not very common if you look at the practical data. On the other hand, if you look at the definition of randomness, it follows that in certain procedures of data analysis there must be, on pure mathematical grounds, some kind of a room for non-uniqueness, because if there were no room for non-uniqueness, you couldn't explain any noise present in the data and yet noise is always present in the data. So if instead of assuming a priori a probabilistic model, you simply analyze the data, but in such a fashion that you allow some possibility of non-uniqueness. You will see that the real randomness in the data will take advantage of this non-uniqueness that you have allowed and you have very nicely separated the random part of your computational results with those which depend on phenomena that are only veiled by the noise, but are fundamentally deterministic.

In other words, it is perfectly possible to avoid the big problem that you cannot deal with probabilistic models, unless you have a procedure for identifying these models. In fact, there is a further belief, perhaps not absolutely final, that in modern data analysis procedures you can identify non-probabilistic deterministic underlying phenomena, but you have very little chance of really analysing the mechanism of random noise, that is somehow superimposed on the data, because even in identifying random noise, there is an additional level of non-uniqueness, and two different models for the noise work just as well. So that there is no way of distinguishing between them, there is no identification of noise in the sense of classical probabilistic models. In other words, the only chance you have of dealing with randomness is to allow non-uniqueness in the right way and then everything is fine.

I just want to make one more comment. Namely there is a famous statement by Einstein on the next slide (10):

10

Einstein:

God does not
play dice

[Gott würfelt nicht]

$\not\Rightarrow$ Nature does not admit
randomness

which everybody has heard of. I think that this statement of Einstein is a very sneaky one. Einstein could have said for example that I don't believe that there is randomness in nature. That everything in nature is deterministic if I look at it the right way. But Einstein did not say that and I think he would have been laughed at, because there is too much in nature that is not uniquely determined. In the sense of microscopic phenomenon, such as the shape of two trees, maybe from practically the same seed and never exactly the same. So there is some randomness in nature. But Einstein didn't say: "I don't believe there is randomness in nature". He merely says that God doesn't play dice.

Now what do you mean, doesn't play dice? Well, I can say that he meant that God (that is to say nature) doesn't act in such a fashion that the classical currently available probability theory can be applied, because this probability theory comes after all from playing dice. It is a good model for playing dice, but it doesn't follow that it is a good model for nature. Or he could have even said: "Well, I don't want to use an overly technical language, but if you ask me I don't believe in the Kolmogorov axioms being a physical law". Well, the Kolmogorov axioms define a certain kind of probabilistic world, which is perfectly okay for playing dice, or roulette, or shuffling cards and that sort of thing. But why should they be relevant to the real world? In other words, although this may be stretching the point a little bit, I believe you could say, and perhaps Einstein would smile approvingly, that the point he really wanted to make is that we don't know. I mean why should we assume that God plays dice like we play dice as an explanation of what happens in nature? We just have to look a little more. We simply don't know. In other words, it is philosophy in the sense of Bertrand Russell.

Thank you very much.