The Chaitin Interview II: Defining Randomness

https://mindmatters.ai/podcast/ep125

Robert J. Marks:

What is randomness? We talk today with the great mathematician and computer scientist Gregory Chaitin about a landmark paper he published when still a teenager.

Announcer:

Welcome to Mind Matters News, where artificial and natural intelligence meet head on. Here's your host, Robert J. Marks.

Robert J. Marks:

Greetings. We are chatting with Gregory Chaitin, the co-founder of the field of algorithmic information theory that explores properties of computer programs. Professor Chaitin, welcome again.

Gregory Chaitin:

Thank you very much, Bob. Yeah, in particular, it looks at the size of computer programs in bits. More technically you ask, what is the size in bits of the smallest computer program you need to calculate a given digital object? That's called the program size complexity or the algorithmic information content.

Robert J. Marks:

I've heard you call those elegant programs.

Gregory Chaitin:

Well, elegant programs are the ... Yeah, that's the smallest program that has the output that it does, and its size in bits will be the measure of complexity or of algorithmic information content.

Robert J. Marks:

Yes, yes. Now, this goes back to your co-founding the area of algorithmic information theory, and as a teenager, you published a landmark paper. The title of it was “On the Length of Programs for Computing Finite Binary Sequences.” You published it in the Journal of the Association for Computing Machinery, which I guess is the oldest journal associated with theoretical computer science. It was founded in 1947, right after World War II. This was a landmark paper in the founding of algorithmic information theory, and you covered a number of topics. One of them, which is just fascinating, is a brand new concept of the idea of randomness. You offered a whole new definition of randomness. Do you have a definition for randomness in general?

Gregory Chaitin:

Well, look, the normal definition of randomness is, if the process that produces something is unpredictable, like tossing a coin, if you have a fair coin and you keep tossing it, that's going to give you a random sequence of heads and tails. That's - you look at how it was generated, the sequence. But what I wanted was a definition that doesn't ... I wanted a mathematical definition, because you see if you toss a coin, you could get all heads, and that isn't random, but it is possible.

Robert J. Marks:

Yes.

Gregory Chaitin:

I didn't like that definition, so I wanted a definition of lack of structure. You see, with the normal coin tosses, actually every possible finite sequence of heads and tails in a sense is equally random, because they were all generated by tossing a fair coin, independent tosses of a fair coin. But some of them, all heads has a lot of structure, all tails have a lot of structure, alternating heads and tails have a lot of structure. I was looking at something that ignored how the sequence is generated and just looked at it and said, is there structure here or isn't there?

Gregory Chaitin:

Now, the reason for doing this is because you can think of a theory, a physicist theory to explain a phenomenon as a program, a software that can calculate the predictions. If the program is short, then you have a very comprehensible theory and a lot of structure. But if the program is the same size in bits as the number of bits of experimental data, then that's not much of an explanation, it's not much of a theory, because there always is a program the same size in bits as the bits of data.

Gregory Chaitin:

Why? Because it just puts the data into the program directly and print it out. That can always be done. But the smaller the program is compared in size in bits to the number of bits of data that you're trying to explain, and I'm talking about an explanation that gives no noise. It's not a statistical theory. It has to give every bit correctly of the data. If that's a small program, then you have a good theory. If you have two theories and one of them is a smaller program than the other, the smaller program is a better theory, if two of them calculate the exact sequence of your experimental data. It's sort of a model of the scientific method.

Gregory Chaitin:

Now, I'm not using equations. Normally people talk about ... There's a lot of talk about complexity in discussions of the philosophy of science, but they're talking about the complexity of the equations, for example. That's very hard to define and make a mathematical theory about it, because mathematical notation changes. But if you have to explain to a computer how to calculate the observations, there are universal term machines. There are optimum computers that give the smallest programs, and that's a good basis for a mathematical theory, a more precise definition of complexity.

Gregory Chaitin:

See, so when I was a kid, I was reading a book by Karl Popper called *The Logic of Scientific Discovery*, I think it was called. He has a whole chapter on simplicity, and he points out some remarks of Herman Weyl on this, another book that I read. Hermann Weyl was a wonderful ... He was a student of Hilbert. He was a wonderful mathematician and mathematical physicist, and he wrote two books on philosophy where he says that the notion of causality of a theory, really saying that something is governed by a scientific law, is meaningless unless you have a notion of complexity, because there's always a law.

Gregory Chaitin:

You can always find ... He points out, this goes back to Leibniz in 1686. You can also always find an equation passing through points of data on a graph, a thing called Lagrangian interpolation, which will produce an algebraic equation that passes through any finite set of points. Leibniz makes a similar remark that Weyl was aware. You have to have a notion of complexity as well as a notion of what a law is, because otherwise it's meaningless to say that there's a theory for something.

Gregory Chaitin:

I think this is a very deep remark. The question was, I think Weyl also says, it's tough to define this precisely because mathematical notation changes. What are you going to use? Are you going to use Bessel functions in your equation, for example? They change as a function of time, so it seems a bit arbitrary. Now, taking a universal computer and looking at the size in bits of a program gives a more definite notion of complexity that you measure in bits.

Gregory Chaitin:

Also, there's a problem, because if you look at what Weyl discussed and what Leibniz discussed, they're talking about points of data that a scientist has on graph paper, and these points are infinite precision information. They're real numbers. In theory, a point is infinite precision, so it's an infinite amount of information. A key step in an algorithm information theory is that you replace the original problem, which was points of data on graph paper and an equation passing through those points, which doesn't work out too well, although it's closer to the real case in real physics.

Gregory Chaitin:

You replace it by discrete and finite amounts of information. You think of the physical scientific data you're trying to explain as a finite sequence of zeros and ones, and then the program, which is your theory, is also a finite sequence of zeros and ones, in binary. Inside computers it's always binary. Then it's very easy to compare how many bits in your theory versus how many bits in your data, you see? The simplest theory is the best, and if there is no theory simpler than the data you are trying to explain, then the data is random. It has no structure, because any sequence of bits, you can calculate it from a program the same size in bits as the data. That doesn't enable you to distinguish between a sequence of bits with structure, from a sequence of bits that has no structure. It's when you say that the program has to be simpler than the data you're trying to explain it. Your theory has to be simpler than the data. Then it's a theory.

Gregory Chaitin:

This idea goes back to the discourse on metaphysics, which is a relatively short text of Leibniz. The original is in French, it's called *Discours de Metaphysique*. It was found nearly a century after Leibniz died among his papers. The person who found it gave it this name, and Weyl, following the Germanic tradition, had studied a lot of philosophy. Leibniz is a hero in the German speaking world. He's less known outside the German speaking world. His two books on philosophy mention Leibniz and mention this idea of Leibniz. Then Popper refers to it in *The Logic of Scientific Discovery*, which was originally in German also, by the way. He was a refugee from the Second World War.

Gregory Chaitin:

Algorithmic Information Theory takes up this question and changes the context, from a theory being an equation passing through a set of points which have an infinite precision, to making everything discrete and have a finite number of bits. Then, mathematically, you're in business. This was the fundamental idea that inspired ... At least it inspired me, to try to work out a detailed theory. I had these papers in high school, but I did many versions of the theory, and the one that I regard as definitive, it's called "A Theory of Program Size, Formally Identical to Information Theory," also published in the journal of the ACM. I think it was 1975 in the ACM journal. "Theory of Program Size, Formerly Identical to Information Theory." The original versions had some problems, and I think the definitive version is from 1975.

Gregory Chaitin:

Now, my interest in this was philosophical. I wanted to understand what a theory is, how to measure its complexity, but mostly I was interested in incompleteness, because it turns out that this notion of complexity asked for the size of the smallest program to calculate something. This is how you measure the algorithmic information content of that digital object. It's a nice definition. You have a nice mathematical theory, but you can never calculate it. Well, except in a finite number of cases for various small programs. Everywhere you go you get incompleteness in this theory, things that you can define but you can't calculate. Incompleteness sort of hits you in the face in this theory, and my main interest was in incompleteness, and trying to extend the work of Gödel and Turing that I had studied as a teenager, on incompleteness. But there are other people who will have more practical interests, and making this criteria that a good theory is a small one, you can apply that to predicting future observations by looking at the size of-

Robert J. Marks:

This is kind of the work of Solomonoff, right?

Gregory Chaitin:

Yeah, he was interested in prediction. I was more interested in looking at a given string of bits and asking, does it have structure or not, and the incompleteness results regarding this question. For example, most strings of bits have no structure according to this definition. They cannot be compressed into a smaller program, but it turns out you can almost never prove it. You can show that it's very high probability, but can only be provable for extremely small sequences. That was what fascinated me, but Solomonoff was interested in artificial intelligence.

Gregory Chaitin:

Marvin Minsky praised a Solomonoff's theory, and about a year before he died, in the world science fair in New York city, Marvin and I were on a panel and we were filmed with Rebecca Goldstein and Mario Livio, and a Nobel Prize winner in biology was the moderator, and it's an hour and a half on film. At the very end, Marvin surprised me by saying that in his view, the decisive step forward from Gödel is using this approach to making predictions. Now, he says it can't be done, it would require an infinite amount of computing to get precisely the best prediction according to this criterion. But he says he suspects there are good approximations, and people have to work on that. In fact, indeed people have worked on that. Hector Zenil has done a lot with using approximate versions of these ideas, which are computable.

Robert J. Marks:

Yes. I'm an engineer. I'm an engineer that loves mathematics, and I teach a graduate course on information theory, including both Shannon and algorithmic information theory. I explain the randomness in this fashion, and let me pass it by you just to make sure that I'm explaining it right. It's the maximal degree to which a sequence of bits can be compressed. We talk about compressing files using Lempel-Ziv and zip files and PNG images, where they compress the image in order to transmit it over a channel. They do that much like dehydrated food. You take the water out, you ship it, because the shipping is cheaper, and then you hydrate it on the other side. But this Lempel-Ziv doesn't ... I've tried it on a number of different images, like for example an image and the scaled image, and it doesn't take. Clearly, the Lempel-Ziv and the zip files that we generate are not the smallest. I make the case that there must be a smallest file that generates the random output, and that is the concept of what you would call elegant programs. Is that pretty accurate?

Gregory Chaitin:

Yeah, of course. That's a very good way to explain it. I was interested in Shannon's theory of information and noiseless coding, the coding with redundancy when there's noise. One of my papers, one of my first papers was in 1970 in the information theory transactions. I didn't start there with the philosophy of science. That interested me, but I didn't think the readers of that article would be very interested, so I started with the Shannon diagram and said, well, let's send the smallest program to calculate something. That's noiseless coding. That's going to be the most compressed version. Then at the other end, what you do is you ... That's a kind of a universal scheme for compressing. That'll give you the best compression possible. If you use a computer at the other end to get back to the original message, you run the program and it gives you the original message.

Gregory Chaitin:

The only problem with this is, you can't get the best program, the most concise, compressed form of the message, according to this definition. It exists, in the Platonic world of mathematics, but actually finding it is impossible, in fact, in general. That's the philosophically interesting part, but you can view it from an engineering point of view. If you're interested in AI and making predictions, and Marvin Minsky was, and Ray Solomonoff was, then this is ... It's a very interesting new approach, and as Minsky said ... Well, he likes to be provocative. He said, everybody should work on this, to find practical approximations to do this impossible task. The person I know who's done it best has been Hector Zenil and his collaborators. He's in Europe.

Gregory Chaitin:

My interests were more proving theorems and in particular proving incompleteness, and the light it sheds on the scientific method. It's metaphysics. What is a theory? What is the simplicity of a theory? What's a good theory? If you have two theories, which will you look at? That was my starting point. But I was reading Shannon, I was reading Turing, I was reading Von Neumann on game theory. Actually, I had forgotten the definition of randomness that I had put in the essay question to get into the Columbia program for bright high school students, and I remembered it when I read a footnote in Von Neumann. His theory for certain games says, the best thing to do is to toss a coin. That's because ... Well, it's a little long to explain. He has a footnote saying, actually, does that mean there's only a theory of games in a world where there exists randomness?

Gregory Chaitin:

I said, no, you can have a theory that ... You see, everyone will know the theory, so if the theory says, toss a coin, the fact that everyone knows the theory doesn't mean that you're dead, because your opponent will know what to do. But another alternative is, instead of tossing a coin, is if it's an uncomputable sequence of moves that you should make. That way the theory could tell you to pick an unstructured sequence, and then it's not a contradiction because you and your opponent will know the theory, but he won't be able to use that knowledge against you. But unfortunately you won't be able to, unless an Oracle or God gives you an unstructured maximum program size complexity sequence, you won't be able to do what the theory says you should do. But, in theory, it shows that in a world without randomness you could also have a theory that would be impractical, but it would tell you what would be the best thing to do.

Gregory Chaitin:

It was that footnote where Von Neumann says, there's a strange aspect of this theory that it seems to depend on quantum mechanics and the fact that the universe contains randomness. I think he says, this could be discussed in greater length, but he leaves it there. I said, oh, the random sequence is the unstructured sequence I had thought of in that answer to that essay question at Columbia University. That would also work. Now, how you get it, I don't know, but it would work. I was playing with it. I remembered the definition and then I started to work out the mathematics. That was the summer between my first year and my second year at City College, and then the Dean excused me from attending classes because I was working on this immense paper.

Robert J. Marks:

That's something.

Gregory Chaitin:

The Dean was a mathematician, by the way, was a professor of mathematics at that time. I was at City College where Emile Post had been. They had his photograph on the wall of the office of the chairman of the math department.

Robert J. Marks:

Your test at Columbia reminds me of a test that I give. If you give a test where all the problems are simple, you get kind of a histogram with a little peak. If you make them all hard, you get another peak on the other end, so an ideal test should have a gradient. I tell the students that there's going to be some simple ones, some medium ones and some hard ones. Sometimes I ask questions which I don't know the answer for, so I tell them, if you get the answer to some of the harder questions, we have a publication. I think that that's kind of what you did at the Columbia entrance test, right?

Gregory Chaitin:

Yeah. Well, that reminds me of a joke of my late friend Jacob Schwartz, a mathematician at Courant Institute. He floated the idea of putting Fermat's last theorem as a ... Including it in the problems in an important exam in mathematics, in the hope that some undergraduate would come up with a wonderful short - Fermat claimed he had a short proof - a wonderful short proof, not knowing that this was a immensely hard problem that many famous mathematicians had worked on unsuccessfully for a long time. But that's not how it was solved. It was solved with a very fine, sophisticated mathematician working on it in secret for years, and it's a very long proof. Wiles's proof.

Robert J. Marks:

Amazing. I guess Fermat was wrong when he said he could fit the proof in the margin.

Gregory Chaitin:

You know, that's an interesting historical question. When Fermat said he had a proof, he always had a proof, I think. The only case that was left hanging and was that. He was a very superb mathematician, Fermat, so I personally think he had a proof, but we haven't figured it out. It's based on different ideas than Wiles's proof, because those concepts didn't exist at that time. But I could be wrong.

Gregory Chaitin:

By the way, there's a lovely musical comedy about all of this called “Fermat's Last Tango,” and it's available on the web. It's a conflict between the ghost of Fermat that doesn't want Wiles to find the proof, and Wiles, and Wiles's wife who would like Wiles to come back to earth because he's working all the time and this in secret, and she doesn't get to see him very much, nor do his children. It's great fun. It's a musical comedy. It's written by someone who knows mathematics, so the jokes are all good math jokes.

Robert J. Marks:

I've got to ask, was it successful? It seems that the audience would be somewhat limited.

Gregory Chaitin:

People were falling off their seats. It was wonderful. They're not terribly sophisticated math jokes, but they're all correct, and the songs are correct, and the history that they give. There's a song where Fermat is taunting Wiles, your proof has got a hole. It's very clever, and they also have Heaven where there are the ghosts of Euclid, Gauss, Pythagoras looking over all of this. And the different styles of music, it's great fun. They're dancing also. It's wonderful.

Robert J. Marks:

I tell you, it, it takes a lot of talent to take something, a mathematical proof, and make it into an entertaining play. That is wonderful.

Gregory Chaitin:

It's a musical comedy, which sounds even harder, right?

Robert J. Marks:

Yes, exactly. We're going to find this. You say it's YouTube?

Gregory Chaitin:

The Clay Foundation, the one that has the Clay prizes for a million dollars for those very hard problems, the Clay Institute, something like that. They paid the money to make a DVD, and then somebody put it on YouTube, so you can now get it like that. I recommend it highly. Oh, and then there's a song taunting Wiles, where Fermat is taunting Wiles again, to try to keep him from finding his proof, saying mathematics is a young man's game, and how old are you?

Robert J. Marks:

Oh, did Fermat do this when he was older?

Gregory Chaitin:

No, this is a fantasy. This is Fermat's ghost. Oh, Fermat? I'm not sure. It's a marginal comment in a copy of Diophantus, and I think the book was even lost. PT Bell thinks somebody stole it. Anyway, go and see this musical comedy. It's great fun, and all the math is right.

Robert J. Marks:

We will find the link to that and post it on the podcast notes. Thank you, Professor Chaitin. We've been talking to Gregory Chaitin, who is the co-founder of algorithmic information theory, on some just fascinating, fascinating things, ranging from what is randomness, to Fermat's last theorem, a musical comedy, Fermat's Last Tango. Until next time, be of good cheer.

Announcer:

This has been Mind Matters News, with your host Robert J. Marks. Explore more at mindmatters.ai. That's mindmatters.ai. Mind Matters News is directed and edited by Austin Egbert. The opinions expressed on this program are solely those of the speakers.