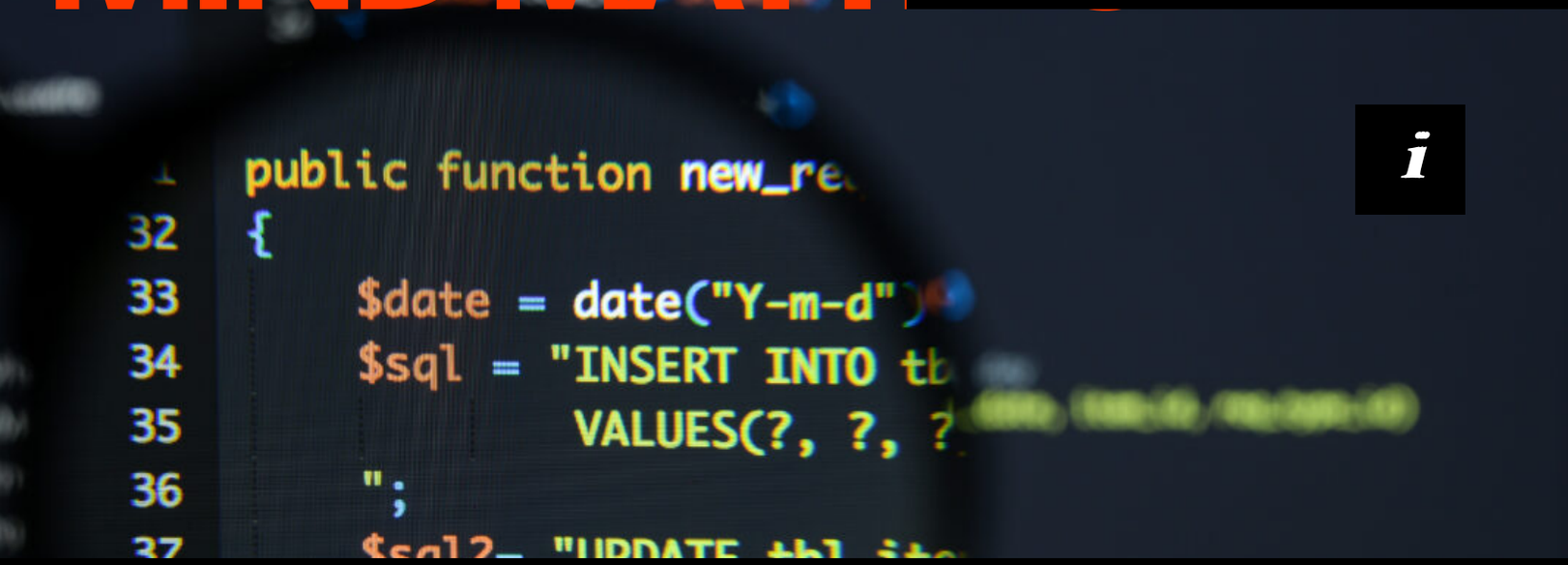


HOW DID RAY SOLOMONOFF KICKSTART ALGORITHMIC INFORMATION THEORY?

He started off the long pursuit of the shortest effective string of information that describes an object

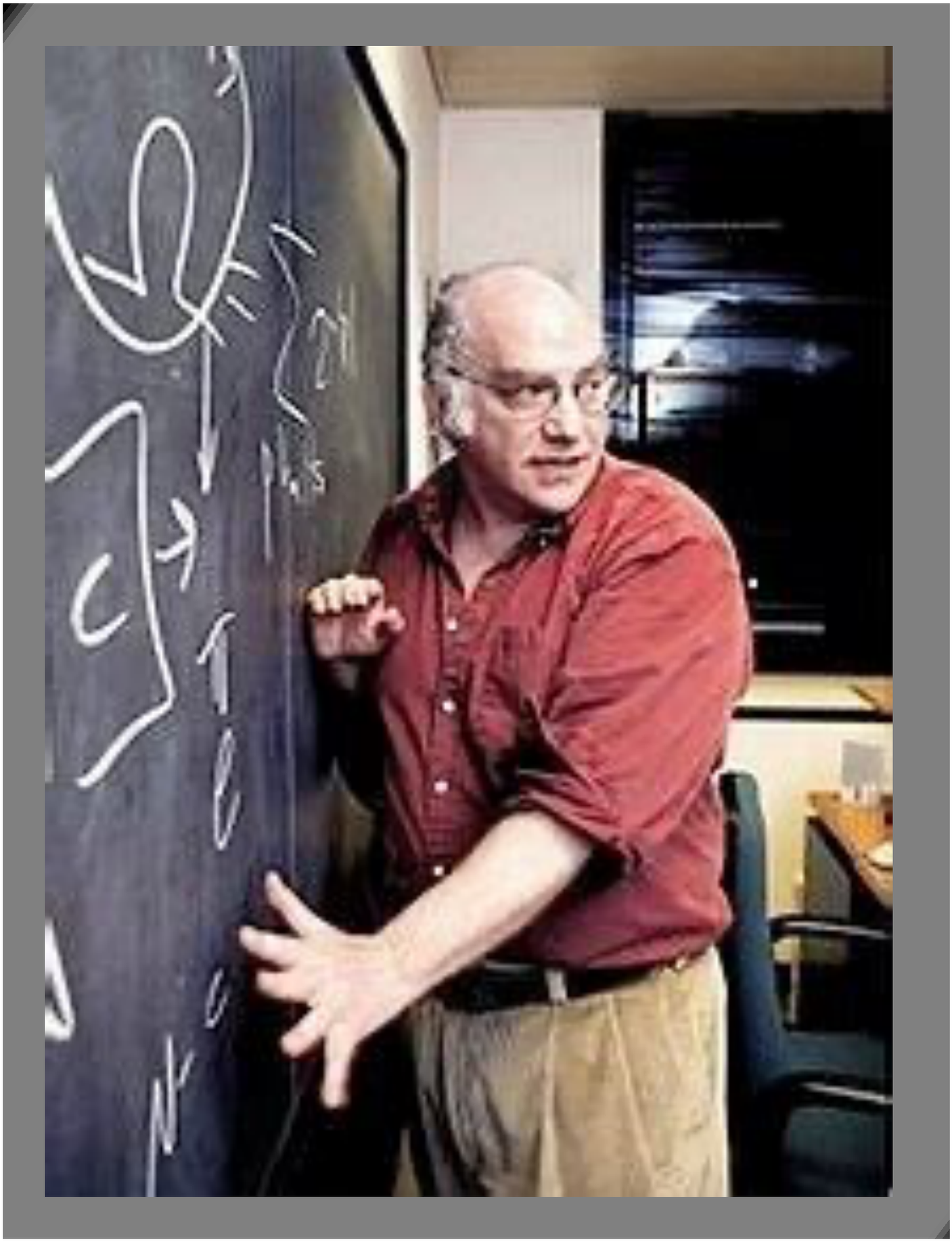
BY **NEWS** ON MARCH 15, 2021 10 MINUTE READ

Share

In last week's podcast,, "The Chaitin Interview II: Defining Randomness," Walter Bradley Center director Robert J. Marks interviewed mathematician and computer scientist Gregory Chaitin on how best to describe true randomness but also on what he recalls of Ray Solomonoff (1926–2009), described in his obit as the "Founding Father of algorithmic information theory."

00:00 00:00

This portion begins at 10:30 min. A partial transcript, Show Notes, and Additional Resources follow.



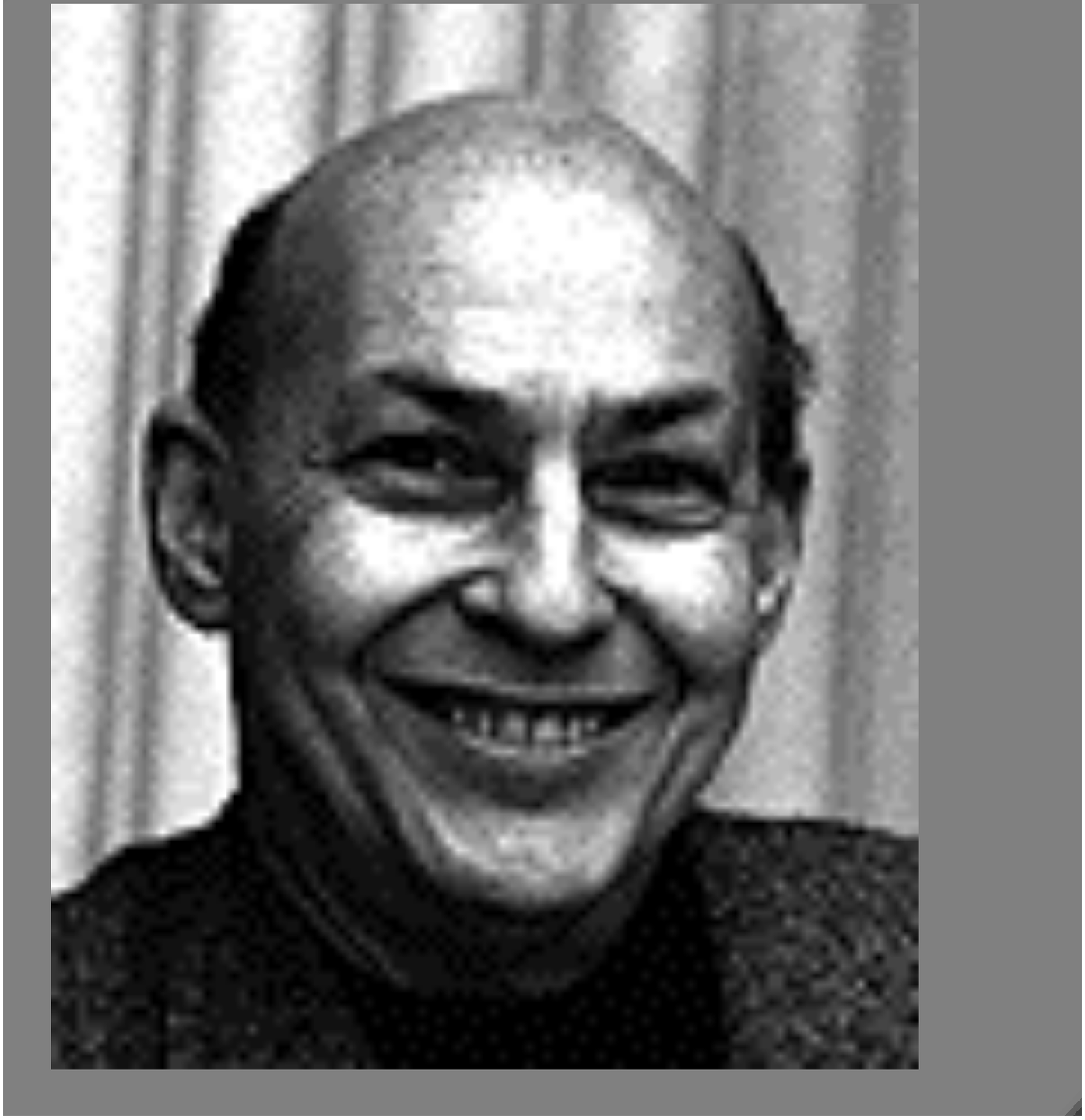
Gregory Chaitin (pictured): Ray Solomonoff was interested in prediction but I was more interested in looking at a

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.

> *Note:* According to his 2009 obituary by Paul Vitanyi, Solomonoff was "the first inventor of Algorithmic Information Theory which deals with the shortest effective description length of objects and is commonly designated by the term 'Kolmogorov complexity.'" Kolmogorov complexity is a measure of meaningful information as opposed to noise.

Gregory Chaitin: Marvin Minsky (pictured) praised Solomonoff's theory. About a year before he died, at the World Science Fair in New York City, Marvin and I were on a panel filmed with Rebecca Goldstein and Mario Livio. A Nobel Prize winner in biology was the moderator. 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 suspects there are good approximations, and people have to work on that. In fact, people have worked on that. Hector Zenil has done a lot with using approximate versions of these ideas, which are computable.

> *Note:* From the MIT obituary notice for Marvin Minsky (1927–2016): "Minsky, a professor emeritus at the MIT Media Lab, was a pioneering thinker and the foremost expert on the theory of artificial intelligence. His 1985 book "The Society of Mind" is considered a seminal exploration of intellectual structure and function, advancing understanding of the diversity of mechanisms interacting in intelligence and thought. Minsky's last book, "The

Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind," was published in 2006."

Also: "Minsky viewed the brain as a machine whose functioning can be studied and replicated in a computer — which would teach us, in turn, to better understand the human brain and higher-level mental functions: How might we endow machines with common sense — the knowledge humans acquire every day through experience? How, for example, do we teach a sophisticated computer that to drag an object on a string, you need to pull, not push — a concept easily mastered by a two-year-old child?" Still working on that.

Robert J. Marks (pictured): 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 first papers was in 1970 in IEEE Transactions in Information Theory. 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.

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 in Europe.

My interests were more in 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?

...

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.

> *Note:* Without randomness, would games even exist? Something other than mere calculation of inevitable results makes them interesting.

Here's Ray Solomonoff:

Next: Fermat's Last Theorem hits Broadway

Previous: Chaitin's discovery of a way of describing true randomness. He found that concepts f rom computer programming worked well because, if the data is not random, the program should be smaller than the data. So, Chaitin on randomness: The simplest theory is best; if no theory is simpler than the data you are trying to explain, then the data is random.

Here are the stories, with links, to an earlier recent podcast discussion with Gregory Chaitin:

Gregory Chaitin's "almost" meeting with Kurt Gödel. This hard-to-find anecdote gives some sense of the encouraging but eccentric math genius. Chaitin recalls, based on this and other episodes, "There was a surreal quality to Gödel and to communicating with Gödel."

Gregory Chaitin on the great mathematicians, East and West: Himself a "game-changer" in mathematics, Chaitin muses on what made the great thinkers stand out. Chaitin discusses the almost supernatural awareness some mathematicians have had of the foundations of our shared reality in the mathematics of the universe.

and

How Kurt Gödel destroyed a popular form of atheism. We don't hear much about logical positivism now but it was very fashionable in the early twentieth century. Gödel's incompleteness theorems showed that we cannot devise a complete set of axioms that accounts for all of reality — bad news for positivist atheism.

You may also wish to read: Things exist that are unknowable: A tutorial on Chaitin's number *(Robert J. Marks)*

and

Five surprising facts about famous scientists we bet you never knew: How about juggling, riding a unicycle, and playing bongo? Or catching criminals or cracking safes? Or believing devoutly in God... *(Robert J. Marks)*

Show Notes

- **00:27** | Introducing Gregory Chaitin
- **01:12** | Chaitin's landmark paper published in his teen years
- **02:04** | Chaitin's definition of randomness
- **08:43** | Metaphysics
- **10:30** | Chaitin's philosophical interest
- **19:24** | Fermat's Last Theorem

Additional Resources

- Gregory Chaitin's Website
- "On the length of programs for computing finite binary sequences," by Gregory Chaitin, published when he as a teenager. (Journal of the ACM (JACM) 13, no. 4 (1966): 547-569).
- *Unravelling Complexity: The Life and Work of Gregory Chaitin*, edited by Shyam Wuppuluri and Francisco Antonio Doria
- Karl Popper
- *The Logic of Scientific Discovery* by Karl Popper
- Hermann Weyl
- Gottfried Wilhelm Leibniz
- Leibniz's *Discours de Métaphysique*

(in English)

- "A Theory of Program Size Formally Identical to Information Theory" by Gregory Chaitin (Journal of the ACM (JACM) 22, no. 3 (1975): 309-440).
- Ray Solomonoff
- Marvin Minsky
- *Fermat's Last Tango* on YouTube

Podcast Transcript Download

(Portions of this transcript have been altered to clarify the content).

MIND MATTERS NEWS

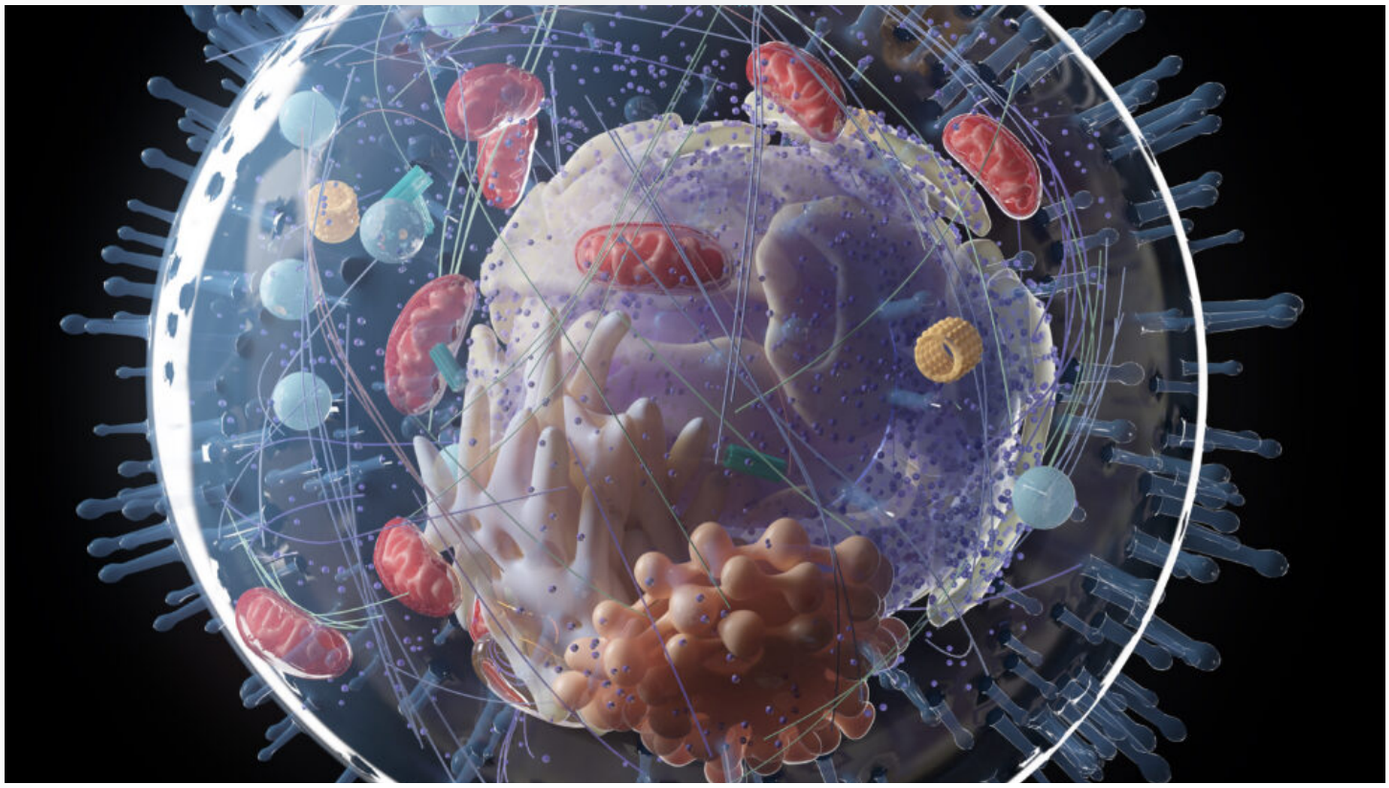
Breaking and noteworthy news from the exciting world of natural and artificial intelligence at MindMatters.ai.

CONTINUE READING

PREVIOUS POST

← THE STARTLING ENERGY COSTS OF BITCOIN

NEXT POST



A NEW THEORY LINKS CONSCIOUSNESS TO BIOELECTRICITY →

About Mind Matters and The Center for Intelligence

Mind Matters features original news and analysis at the intersection of artificial and natural intelligence. Through articles and podcasts, it explores issues, challenges, and controversies relating to human and artificial intelligence from a perspective that values the unique capabilities of human beings. *Mind Matters* is published by the Walter Bradley Center for Natural and Artificial Intelligence.