Lex Fridman Podcast #428 - Sean Carroll: General Relativity, Quantum Mechanics, Black Holes & Aliens

Published - April 22, 2024

Transcribed by - thepodtranscripts.com

Lex Fridman
The following is a conversation with Sean Carroll. His third time in this podcast. He is a theoretical physicist at Johns Hopkins, host of the Mindscape Podcast that I personally love and highly recommend, and author of many books, including the most recent book series called The Biggest Ideas in the Universe. The first book of which is titled Space, Time, and Motion. It's on the topic of general relativity. The second coming out on May 14th, you should definitely pre-order it, it's titled the Quanta and Fields. That one is on the topic of quantum mechanics. Sean is a legit, active, theoretical physicist and at the same time is one of the greatest communicators of physics ever. I highly encourage you listen to his podcast, read his books, and pre-order the new book to support his work. This was, as always, a big honor and a pleasure for me. This is Lex Fridman Podcast. To support it, please check out our sponsors in the description. Now, dear friends here's Sean Carroll. In book one of the series, The Biggest Ideas in the Universe called Space, Time, Motion, you take on classical mechanics, general relativity by taking on the main equation of general relativity and making it accessibly easy to understand. Maybe at the high level, what is general relativity? What's a good way to start to try to explain it?

Sean Carroll
Probably the best way to start to try to explain it is special relativity, which came first, 1905. It was the culmination of many decades of people putting things together. But it was Einstein in 1905. In fact, it wasn't even Einstein. I should give more credit to Minkowski in 1907. Einstein in 1905 figured out that you could get rid of the ether, the idea of a rest frame for the universe and all the equations of physics would make sense with the speed of light being a maximum. But then it was Minkowski who used to be Einstein's professor in 1907 who realized the most elegant way of thinking about this idea of Einstein's was to blend space and time together into spacetime to really imagine that there is no hard and fast division of the four-dimensional world in which we live into space and time separately. Einstein was at first dismissive of this. He thought it was just like, "Oh, the mathematicians or over-formalizing again." But then he later realized that if spacetime is a thing, it can have properties and in particular it can have a geometry. It can be curved from place to place. That was what let him solve the problem of gravity. He had previously been trying to fit in what we knew about gravity from Newtonian mechanics, the inverse square law of gravity, to his new relativistic theory. It didn't work. The final leap was to say gravity is the curvature of spacetime, and that statement is basically general relativity.

Lex Fridman
The tension with Minkowski was he was a mathematician.

Sean Carroll
Yes.

Lex Fridman
It's the tension between physics and mathematics. In fact, in your lecture about this equation, one of them, you say that Einstein is a better physicist than he gets credit for.

Sean Carroll
Yep. I know that's hard. That's a little bit of a joke there, right?

Lex Fridman
Yeah.

Sean Carroll
Because we all give Einstein a lot of credit. But then we also, partly based on fact, but partly to make ourselves feel better, tell ourselves a story about how later in life, Einstein couldn't keep up. There were younger people doing quantum mechanics and quantum field theory and particle physics, and he was just unable to really philosophically get over his objections to that. I think that that story about the latter part is completely wrong, almost 180 degrees wrong. I think that Einstein understood quantum mechanics as well as anyone, at least up through the 1930s. I think that his philosophical objections to it are correct. He should actually have been taken much more seriously about that. What he did, what he achieved in trying to think these problems through is to really basically understand the idea of quantum entanglement, which is important these days when it comes to understanding quantum mechanics. Now, it's true that in the '40s and '50s he placed his efforts in hopes for unifying electricity and magnetism with gravity. That didn't really work out very well. All of us try things that don't work out. I don't hold that against him. But in terms of IQ points, in terms of trying to be a clear-thinking physicist, he was really, really great.

Lex Fridman
What does greatness look like for a physicist? How difficult is it to take the leap from special relativity to general relativity? How difficult is it to imagine that, to consider spacetime together and to imagine that there's a curvature to this whole thing?

Sean Carroll
Yeah. That's a great question. I think that if you want to make the case for Einstein's greatness, which is not hard to do, there's two things you point at. One is in 1905, his famous miracle year, he writes three different papers on three wildly different subjects, all of which would make you famous just for writing that one paper. Special relativity is one of them. Brownian motion is another one, which is just the little vibrations of tiny little dust specks in the air. But who cares about that? What matters is it proves the existence of atoms. He explains Brownian motion by imagining there are molecules in the air and deriving their properties. Brilliant. Then he basically starts the world on the road to quantum mechanics with his paper on, which again, is given a boring label of the photoelectric effect. What it really was is he invented photons. He showed that light should be thought of as particles as

well as waves. He did all three of those very different things in one year. Okay. But the other thing that gets him genius status is, like you say, general relativity. This takes 10 years from 1905 to 1915. He wasn't only doing general relativity. He was working on other things. He invented refrigerator. He did various interesting things. He wasn't even the only one working on the problem. There were other people who suggested relativistic theories of gravity. But he really applied himself to it. I think as your question suggests, the solution was not a matter of turning a crank. It was something fundamentally creative. In his own telling of the story, his greatest moment, his happiest moment was when he realized that if the way that we would modern ... say it in modern terms, if you were in a rocket ship accelerating at 1G, at acceleration due to gravity, if the rocket ship were very quiet, you wouldn't be able to know the difference between being in a rocket ship and being on the surface of the earth. Gravity is not detectable or at least not distinguishable from acceleration. Number one, that's a pretty clever thing to think. But number two, if you or I had had that thought, we would've gone, "Huh. We're pretty clever." He reasons from there to say, "Okay. If gravity is not detectable, then it can't be like an ordinary force." The electromagnetic force is detectable. We can put charged particles around. Positively charged particles and negatively charged particles respond differently to an electric field or to a magnetic field. He realizes that what his thought experiment showed, or at least suggested, is that gravity isn't like that. Everything responds in the same way to gravity. How could that be the case? Then this other leap he makes is, "Oh, it's because it's the curvature of spacetime." It's a feature of spacetime. It's not a force on top of it. The feature that it is, is curvature. Then finally he says, "Okay. Clearly, I'm going to need the mathematical tools necessary to describe curvature. I don't know them, so I will learn them." They didn't have MOOCs or AI helpers back in those days. He had to sit down and read the math papers, and he taught himself differential geometry and invented general relativity.

Lex Fridman
What about the step of including time as just another dimension, combining space and time, is that a simple mathematical leap as Minkowski suggested?

Sean Carroll
It's certainly not simple, actually. It's a profound insight. That's why I said I think we should give Minkowski more credit than we do. He's the one who really put the finishing touches on special relativity. Again, many people had talked about how things change when you move close to the speed of light, what Maxwell's equations of electromagnetism predict and so forth, what their symmetries are. People like Lorenz and Fitzgerald and Poincare, there's a story that goes there. In the usual telling Einstein puts the capstone on it. He's the one who says, "All of this makes much more sense if there just is no ether. It is undetectable. We don't know how fast. Everything is relative." Thus, the name relativity. But he didn't take the actual final step, which was to realize that the underlying structure that he had invented is best thought of as unifying space and time together. I honestly don't know what was going through Minkowski's mind when he thought that. I'm not sure if he was so mathematically

adept that it was just clear to him or he was really struggling it and he did trial and error for a while. I'm not sure.

Lex Fridman
Do you, for him or Einstein, visualize the four-dimensional space, try to play with the idea of time is just another dimension?

Sean Carroll
Oh, yeah. All the time. I mean, we, of course, make our lives easy by ignoring two of the dimensions of space. Instead of four-dimensional spacetime, we just draw pictures of one dimension of space, one dimension of time. The so-called spacetime diagram. I mean, maybe this is lurking underneath your question. But even the best physicists will draw a vertical axis and a horizontal axis and will go space, time. But deep down that's wrong, because you're sort of preferring one direction of space and one direction of time. It's really the whole two-dimensional thing that is spacetime. The more legitimate thing to draw on that picture are rays of light, are light cones. From every point, there is a fixed direction at which the speed of light would represent. That is actually inherent in the structure. The division into space and time is something that's easy for us human beings.

Lex Fridman
What is the difference between space and time from the perspective of general relativity?

Sean Carroll
It's the difference between X and Y when you draw axes on a piece of paper.

Lex Fridman
There's really no difference?

Sean Carroll
There is almost no difference. There's one difference that is important, which is the following; If you have a curve in space, I'm going to draw it horizontally, because that's usually what we do in spacetime diagrams, if you have a curve in space, you've heard the motto before that the shortest distance between two points is a straight line. If you have a curve in time, which is by the way, literally all of our lives, we all evolve in time. You can start with one event in spacetime, and another event in spacetime. What Minkowski points out is that the time you measure along your trajectory in the universe is precisely analogous to the distance you travel on a curve through space. By precisely, I mean it is also true that the actual distance you travel through depends on your path. You can go a straight line, shortest distance and curvy line would be longer. The time you measure in spacetime, the literal time that takes off on your clock also depends on your path, but it depends on it the other way. That the longest time between two points is a straight line. If you zig back and forth in spacetime, you take less and less time to go from point A to point B.

Lex Fridman
How do we make sense of that, the difference between the observed reality and the objective reality are underneath it, or is objective reality a silly notion given general relativity?

Sean Carroll
I'm a huge believer in objective reality. I think that objective reality, objectivity ...

Lex Fridman
You're fan.

Sean Carroll
... is real. But I do think that people are a little overly casual about the relationship between what we observe and objective reality in the following sense. Of course, in order to explain the world, our starting point and our ending point is our observations, our experimental input, the phenomena we experience and see around us in the world. But in between, there's a theory, there's a mathematical formalization of our ideas about what is going on. If a theory fits the data and is very simple and makes sense in its own terms, then we say that the theory is right. That means that we should attribute some reality to the entities that play an important role in that theory, at least provisionally until we can come up with a better theory down the road.

Lex Fridman
I think a nice way to test the difference between objective reality and the observed reality is what happens at the edge of the horizon of a black hole. Technically, as you get closer to that horizon, time stands still?

Sean Carroll
Yes and no. It depends on exactly how careful we are being. Here is a bunch of things I think are correct. If you imagine there is a black hole, spacetime, the whole solution Einstein's equation, and you treat you and me as what we call test particles. We don't have any gravitational fields ourselves. We just move around in the gravitational field. That's obviously an approximation. Okay. But let's imagine that. You stand outside the black hole and I fall in. As I'm falling in, I'm waving to you because I'm going into the black hole, you will see me move more and more slowly. Also, the light for me is redshifted. I kind of look embarrassed, because I'm falling into a black hole. There is a limit. There's a last moment that light will be emitted from me, from your perspective forever. Okay. Now you don't literally see it because I'm emitting photons more and more slowly because from your point of view. It's not like I'm equally bright. I basically fade from view in that picture. Okay. That's one approximation. The other approximation is I do have a gravitational field of my own, and therefore as I approach the black hole, the black hole doesn't just sit there and let me pass

through. It moves out to eat me up because its net energy mass is going to be mine, plus its. But roughly speaking, yes, I think so. I don't like to go to the dramatic extremes because that's where the approximations break down. But if you see something falling into a black hole, you see its clock ticking more and more slowly.

Lex Fridman
How do we know it fell in?

Sean Carroll
We don't. I mean, how would we. Because it's always possible that right at the last minute it had a change of heart and starts accelerating away. If you don't see it passing, you don't know. Let's point out that as smart as Einstein was, he never figured out black holes, and he could have. It's embarrassing. It took decades for people thinking about general relativity to understand that there are such things as black holes. Because basically Einstein comes up with general relativity in 1915. Two years later, Schwarzschild, Karl Schwarzschild derives the solution to Einstein's equation that represents a black hole, the Schwarzschild solution. No one recognized it for what it was until the '50s, David Finkelstein and other people. That's just one of these examples of physicists not being as clever as they should have been.

Lex Fridman
Well, that's the singularity. That's the edge of the theory. The limit. It's understandable that it's difficult to imagine the limit of things.

Sean Carroll
It is absolutely hard to imagine. A black hole is very different to many ways from what we're used to. On the other hand, I mean the real reason, of course, is that between 1915 and 1955, there's a bunch of other things that are really interesting going on in physics. All of particle physics and quantum field theory. Many of the greatest minds were focused on that. But still, if the universe hands you a solution to general relativity in terms of curved spacetime and its mysterious certain features of it, I would put some effort in trying to figure it out.

Lex Fridman
How does a black hole work? Put yourself in the shoes of Einstein and take general relativity to its natural conclusion about these massive things.

Sean Carroll
It's best to think of a black hole as not an object so much as a region of spacetime. Okay. It's a region with the property, at least in classical general relativity, quantum mechanics makes everything harder. But let's imagine we're being classical for the moment. It's a region of spacetime with the property that if you enter, you can't leave. Literally the equivalent of escaping a black hole would be moving faster than the speed of light. They're both precisely equally difficult. You would have to move faster than the speed of light to

escape from the black hole. Once you're in, that's fine. In principle, you don't even notice when you cross the event horizon, as we call it. The event horizon is that point of no return, where once you're inside, you can't leave. But meanwhile, the spacetime is collapsing around you to ultimately a singularity in your future, which means that the gravitational forces are so strong, they tear your body apart and you will die in a finite amount of time. The time it takes, if the black hole is about the mass of the sun to go from the event horizon to the singularity takes about 1 millionth of a second.

Lex Fridman
What happens to you if you fall into the black hole? If we think of an object as information, that information gets destroyed.

Sean Carroll
Well, you've raised a crucially difficult point. That's why I keep needing to distinguish between black holes according to Einstein's theory, General Relativity, which is book one of Spacetime and Geometry, which is perfectly classical. Then come the 1970s, we start asking about quantum mechanics and what happens in quantum mechanics. According to classical general relativity, the information that makes up you when you fall into the black hole is lost to the outside world. It's there, it's inside the black hole, but we can't get it anymore. In the 1970s, Stephen Hawking comes along and points out that black holes radiate. They give off photons and other particles to the universe around them. As they radiate, they lose mass, and eventually they evaporate, they disappear. Once that happens, I can no longer say the information about you or a book that I threw in the black hole or whatever is still there, is hidden behind the black hole because the black hole has gone away. Either that information is destroyed, like you said, or it is somehow transferred to the radiation that is coming out to the Hawking radiation. The large majority of people who think about this belief that the information is somehow transferred to the radiation and information is conserved. That is a feature both of general relativity by itself and of quantum mechanics by itself. When you put them together, that should still be a feature. We don't know that for sure. There are people who have doubted it, including Stephen Hawking for a long time. But that's what most people think. What we're trying to do now in a topic which has generated many, many hundreds of papers called the Black Hole Information Loss Puzzle is figure out how to get the information from you or the book into the radiation that is escaping the black hole.

Lex Fridman
Is there any way to observe Hawking radiation to a degree where you can start getting insight? Or is this all just in the space of theory right now?

Sean Carroll
Right now, we are nowhere close to observing Hawking radiation. Here's the sad fact. The larger the black hole is, the lower its temperature is. A small black hole, like a microscopically small black hole might be very visible. It's given off light. But something like the black hole, the center of our galaxy, 3 million times the mass of the sun or something like that, Sagittarius A star, that is so cold and low temperature that it's radiation will never be observable. Black holes are hard to make. We don't have any nearby. The ones we have out there in the universe are very, very faint. There's no immediate hope for detecting Hawking radiation.

Lex Fridman
Allegedly. We don't have any nearby?

Sean Carroll
As far as we know, we don't have any nearby.

Lex Fridman
Tiny ones be hard to detect somewhere at the edges of the solar system, maybe?

Sean Carroll
You don't want them to be too tiny or they're exploding. They're very bright and then they'll be visible. But there's an absolutely regime where black holes are large enough not to be visible because the larger ones are fainter. Not giving off radiation, but small enough to not been detected through their gravitational effect. Yeah.

Lex Fridman
Psychologically, just emotionally, how do you feel about black holes? They scare you.

Sean Carroll
I love them. I love black holes. But the universe weirdly makes it hard to make a black hole, because you really need to squeeze an enormous amount of matter and energy into a very, very small region of space. We know how to make stellar black holes. A supermassive star can collapse to make a black hole. We know we also have these supermassive black holes, the center of galaxies. We're a little unclear where they came from. I mean, maybe stellar black holes that got together and combined. But that's one of the exciting things about new data from the James Webb Space Telescope is that quite large black holes seem to exist relatively early in the history of the universe. It was already difficult to figure out where they came from. Now it's an even tougher puzzle.

Lex Fridman
These supermassive black holes are formed somewhere early on in the universe. I mean, that's a feature, not a bug, that we don't have too many of them. Otherwise, we wouldn't have the time or the space to form the little pockets of complexity that we'll call humans.

Sean Carroll
I think that's fair. Yeah. It's always interesting when something is difficult, but happens anyway. I mean, the probability of making a black hole could have been zero. It could have been one. But it's this interesting number in between, which is fun.

Lex Fridman
Are there more intelligent alien civilization than there are supermassive black holes?

Sean Carroll
Yeah. I have no idea. But I think your intuition is right that it would've been easy for there to be lots of civilizations then we would've noticed them already and we haven't. Absolutely the simplest explanation for why we haven't is that they're not there.

Lex Fridman
Yeah. I just think it's so easy to make them though. There must be ... I understand that's the simplest explanation. But also ...

Sean Carroll
How easy is it to make life or eukaryotic life or multicellular life?

Lex Fridman
It seems like life finds a way. Intelligent alien civilizations, sure, maybe there is somewhere along that chain a really, really hard leap. But once you start life, once you get the origin of life, it seems like life just finds a way everywhere in every condition. It just figures it out.

Sean Carroll
I mean, I get it. I get exactly what you're thinking. I think is a perfectly reasonable attitude to have before you confront the data. I would not have expected earth to be special in any way. I would've expected there to be plenty of very noticeable extraterrestrial civilizations out there. But even if life finds a way, even if we buy everything you say, how long does it take for life to find a way? What if it typically takes 100 billion years, then we'd be alone.

Lex Fridman
It's a time thing. To you, really most likely, there's no alien civilizations out there. I can't see it. I believe there's a ton of them, and there's another explanation why we can't see them.

Sean Carroll
I don't believe that very strongly. Look, I'm not going to place a lot of bets here. I'm both pretty up in the air about whether or not life itself is all over the place. It's possible when we visit other worlds, other solar systems, there's very tiny microscopic life ubiquitous, but none of it has reached some complex form. It's also possible there isn't any. It's also possible that there are intelligent civilizations that have better things to do than knock on our doors. I think we should be very humble about these things we know so little about.

Lex Fridman
It's also possible there's a great filter where there's something fundamental about once the civilization develops complex enough technology, that technology is more statistically likely to destroy everybody versus to continue being creative.

Sean Carroll
That is absolutely possible. I'm actually putting less credence on that one just because you need to happen every single time. If even one, I mean, this goes back to John von Neumann pointed out that you don't need to send the aliens around the galaxy. You can build self-reproducing probes and send them around the galaxy. You might think, "Well, the galaxy is very big." It's really not. It's some tens of thousands of light years across and billions of years old. You don't need to move at a high fraction of the speed of light to fill the galaxy.

Lex Fridman
If you were an intelligent alien civilization, the dictator of one, you would just send out a lot of probes, self-replicating probes ...

Sean Carroll
100%.

Lex Fridman
... to spread out.

Sean Carroll
Yes. What you should do ... If you want the optimistic spin, here's the optimistic spin. People looking for intelligent life elsewhere often tune in with their radio telescopes, at least we did before Arecibo was decommissioned. That's not a very promising way to find intelligent life elsewhere, because why in the world would a super intelligent alien civilization waste all of its energy by beaming it in random directions into the sky? For one thing, it just passes you by. If we are here on earth, we've only been listening to radio waves for or a couple 100 years. Okay. If an intelligent alien civilization exists for a billion years, they have to pinpoint exactly the right time to send us this signal. It is much, much more efficient to send probes and to park, to go to the other solar systems, just sit there and wait for an intelligent civilization to arise in that solar system. This is the 2001 monolith hypothesis. I would be less surprised to

find a quiescent alien artifact in our solar system than I would to catch a radio signal from an intelligent civilization.

Lex Fridman
You're a sucker for in-person conversations versus remote.

Sean Carroll
I just want to integrate over time. A probe can just sit there and wait, whereas a radio wave goes right by you.

Lex Fridman
How hard is it for an alien civilization, again, you have the dictator of one, to figure out a probe that is most likely to find a common language with whatever it finds.

Sean Carroll
Couldn't I be like the elected leader of alien civilization?

Lex Fridman
Elected leader, democratic leader. Elected leader of a democratic alien civilization. Yes.

Sean Carroll
I think we would figure out that language thing pretty quickly. I mean, maybe not as quickly as we do when different human tribes find each other, because obviously there's a lot of commonalities in humanity. But there is logic in math, and there is the physical world. You can point to a rock and go "rock." I don't think it would take that long. I know that Arrival, the movie, based on a Ted Chiang story suggested that the way that aliens communicate is going to be fundamentally different. But also, they had recognition and other things I don't believe in. I think that if we actually find aliens, that will not be our long-term problem.

Lex Fridman
There's a folks ... One of the places you're affiliated with is Santa Fe, and they approach the question of complexity in many different ways and ask the question in many different ways of what is life, thinking broadly? To you would be able to find it. You'll think you show up, a probe shows up to a planet, we'll see a thing and be like, "Yeah. That's a living thing."

Sean Carroll
Well, again, if it's intelligent and technologically advanced, the more short-term question of if we get some spectroscopic data from an exoplanet, so we know a little bit about what is in its atmosphere, how can we judge whether or not that atmosphere is giving us a signature of life existing? That's a very hard question that people are debating about. I mean, one very simple-minded, but perhaps interesting approach is to say, "Small

molecules don't tell you anything, because even if life could make them something else could also make them. But long molecules, that's the thing that life would produce."

Lex Fridman
Signs of complexity. I don't know. I just have this nervous feeling that we won't be able to detect. We'll show up to a planet. There have a bunch of liquid on it. We take a swim in the liquid. We won't be able to see the intelligence in it, whether that intelligence looks like something like ants or ... We'll see movement, perhaps, strange movement. But we won't be able to see the intelligence in it or communicate with it. I guess if we have nearly infinite amount of time to play with different ideas, we might be able to.

Sean Carroll
I think I'm in favor of this kind of humility, this intellectual humility that we won't know because we should be prepared for surprises. But I do always keep coming back to the idea that we all live in the same physical universe. Well, let's put it this way. The development of our intelligence has certainly been connected to our ability to manipulate the physical world around us. I would guess, without 100% credence by any means, but my guess would be that any advanced kind of life would also have that capability. Both dolphins and octopuses are potential counterexamples to that. But I think in the details, there would be enough similarities that we would recognize it.

Lex Fridman
I don't know how we got on this-

Sean Carroll
... would be enough similarities that we would recognize it.

Lex Fridman
I don't know how we got on this topic, but I think it was from super massive black holes. So if we return to black holes and talk about the holographic principle more broadly, you have a recent paper on the topic. You've been thinking about the topic in terms of rigorous research perspective and just as a popular book writer?

Sean Carroll
Mmhmm.

Lex Fridman
So what is the holographic principle?

Sean Carroll
Well, it goes back to this question that we were talking about with the information and how it gets out. In quantum mechanics, certainly, arguably, even before quantum mechanics

comes along in classical statistical mechanics, there's a relationship between information and entropy. Entropy is my favorite thing to talk about that I've written books about and will continue to write books about. So Hawking tells us that black holes have entropy, and it's a finite amount of entropy. It's not an infinite amount. But the belief is, and now we're already getting quite speculative, the belief is that the entropy of a black hole is the largest amount of entropy that you can have in a region of space-time. It's the most densely packed that entropy can be. What that means is there's a maximum amount of information that you can fit into that region of space, and you call it a black hole. Interestingly, you might expect if I have a box and I'm going to put information in it and I don't tell you how I'm going to put the information in, but I ask, "How does the information I can put in scale with the size of the box?" You might think, "Well, it goes as the volume of the box because the information takes up some volume, and I can only fit in a certain amount." That is what you might guess for the black hole, but it's not what the answer is. The answer is that the maximum information as reflected in the black hole entropy scales as the area of the black hole's event horizon, not the volume inside. So people thought about that in both deep and superficial ways for a long time, and they proposed what we now call the holographic principle, that the way that space-time and quantum gravity convey information or hold information is not different bits or qubits for quantum information at every point in space-time. It is something holographic, which means it's embedded in or located in or can be thought of as pertaining to one dimension less of the three dimensions of space that we live in. So in the case of the black hole, the event horizon is two-dimensional, embedded in a three-dimensional universe. The holographic principle would say all of the information contained in the black hole can be thought of as living on the event horizon rather than in the interior of the black hole. I need to say one more thing about that, which is that this was an idea, the idea I just told you was the original holographic principle put forward by people like Gerard 't Hooft and Leonard Susskind, the super famous physicist. Leonard Susskind was on my podcast and gave a great talk. He's very good at explaining these things.

Lex Fridman
Mindscape Podcast-

Sean Carroll
Mindscape Podcast.

Lex Fridman
Everybody should listen.

Sean Carroll
That's right, yes.

Lex Fridman
You don't just have physicists on.

Sean Carroll
I don't.

Lex Fridman
I love Mindscape.

Sean Carroll
Oh, thank you very much.

Lex Fridman
Curiosity-driven-

Sean Carroll
Yeah, ideas-

Lex Fridman
... exploration of ideas.

Sean Carroll
Fresh ideas from smart people.

Lex Fridman
Yeah.

Sean Carroll
Yeah.

Lex Fridman
But anyway, what I was trying to get at with Susskind and also at 't Hooft were a little vague. They were a little hand wavy about holography and what it meant, where holography, the idea that information is encoded on a boundary really came into its own was with Juan Maldacena in the 1990s and the AdS-CFT correspondence, which we don't have to get into that into any detail, but it's a whole full-blown theory of... It's two different theories. One theory in N dimensions of space-time without gravity, and another theory in N+1 dimensions of space-time with gravity. The idea is that this N dimensional theory is casting a hologram into the N+1 dimensional universe to make it look like it has gravity. That's holography with a vengeance, and that's an enormous source of interest for theoretical physicists these days. How should we picture what impact that has, the fact that you can store all the information you can think of as all the information that goes into a black hole can be stored at the event horizon?

Sean Carroll
Yeah, it's a good question. One of the things that quantum field theory indirectly suggests is that there's not that much information in you and me compared to the volume of space-time we take up. As far as quantum field theory is concerned, you and I are mostly empty space, and so we are not information dense. The density of information in us or in a book or a CD or whatever, computer RAM, is indeed encoded by volume. There's different bits located at different points in space, but that density of information is super-duper low. So we are just like the speed of light or just the big bang for the information in a black hole, we are far away in our everyday experience from the regime where these questions become relevant. So it's very far away from our intuition. We don't really know how to think about these things. We can do the math, but we don't feel it in our bones.

Lex Fridman
So you can just write off that weird stuff happens in a black hole.

Sean Carroll
Well, we'd like to do better, but we're trying. That's why we have an information loss puzzle because we haven't completely solved it. So here, just one thing to keep in mind. Once space-time becomes flexible, which it does according to general relativity and you have quantum mechanics, which has fluctuations in virtual particles and things like that, the very idea of a location in space-time becomes a little bit fuzzy, 'cause it's flexible and quantum mechanics says you can even pin it down. So information can propagate in ways that you might not have expected, and that's easy to say and it's true, but we haven't yet come up with the right way to talk about it that is perfectly rigorous.

Lex Fridman
It's crazy how dense with information a black hole is, and then plus like quantum mechanics starts to come into play, so you almost want to romanticize the interesting computation type things that are going on inside the black hole.

Sean Carroll
You do. You do, but I'll point out one other thing. It's information dense, but it's also very, very high entropy. So a black hole is kind of like a very, very, very specific random number. It takes a lot of digits to specify it, but the digits don't tell you anything. They don't give you anything useful to work on, so it takes a lot of information, but it's not of a form that we can learn a lot from.

Lex Fridman
But hypothetically, I guess as you mentioned, the information might be preserved. The information that goes into a black hole, it doesn't get destroyed. So what does that mean when the entropy is really high?

Sean Carroll
Well, I said that the black hole is the highest density of information, but it's not the highest amount of information because the black hole can evaporate. When it evaporates and people have done the equations for this, when it evaporates, the entropy that it turns into is actually higher than the entropy of the black hole was, which is good because entropy is supposed to go up, but it's much more dilute. It's spread across a huge volume of space-time. So in principle, all that you made the black hole out of, the information that it took is still there, we think, in that information, but it's scattered to the four winds.

Lex Fridman
We just talked about the event horizon of a black hole. What's on the inside? What's at the center of it?

Sean Carroll
No one's been there, so-

Lex Fridman
And came back to tell?

Sean Carroll
... again, this is a theoretical prediction. But I'll say one super crucial feature of the black holes that we know and love, the kind that Schwarzschild first invented, there's a singularity, but it's not at the middle of the black hole. Remember space and time are parts of one unified space-time, the location of the singularity in the black hole is not the middle of space, but our future. It is a moment of time. It is like a big crunch. The big bang was an expansion from a singularity in the past. Big crunch probably doesn't exist, but if it did, it would be a collapse to a singularity in the future. That's what the interiors of black holes are like. You can be fine in the interior, but things are becoming more and more crowded. Space-time is becoming more and more warped, and eventually you hit a limit, and that's the singularity in your future.

Lex Fridman
I wonder what time is on the inside of a black hole.

Sean Carroll
Time always ticks by at one second per second. That's all it can ever do. Time can tick by differently for different people, and so you have things like the twin paradox where two people initially are the same age, one goes off in the speed of light and comes back, now they're not. You can even work out that the one who goes out and comes back will be younger because they did not take the shortest distance path. But locally, as far as you and your wristwatch are concerned, time is not funny. Your neurological signals in your brain

and your heartbeat and your wristwatch, whatever's happening to them is happening to all of them at the same time. So time always seems to be ticking along at the same rate.

Lex Fridman
Well, if you fall into a black hole and then I'm an observer just watching it, and then you come out once it evaporates a million years later, I guess you'd be exactly the same age? Have you aged at all?

Sean Carroll
You would be converted into photons. You would not be you anymore.

Lex Fridman
Right. So it's not at all possible that information is preserved exactly as it went in.

Sean Carroll
It depends on what you might preserve. It's there in the microscopic configuration of the universe. It's exactly as if I took a regular book, made it paper and I burned it. The laws of physics say that all the information in the book is still there in the heat and light and ashes. You're never going to get it. It's a matter of practice, but in principle, it's still there.

Lex Fridman
But what about the age of things from the observer perspective, from outside the black hole?

Sean Carroll
From outside the black hole, doesn't matter 'cause they're inside the black hole.

Lex Fridman
No. Okay. There's no way to escape the black hole-

Sean Carroll
Right. ... except-

Lex Fridman
To let it evaporate. ... to let it evaporate. But also, by the way, just in relativity, special relativity, forget about general relativity, it's enormously tempting to say, "Okay, here's what's happening to me right now. I want to know what's happening far away right now." The whole point of relativity is to say there's no such thing as right now when you're far away, and that is doubly true for what's inside a black hole. So you're tempted to say, "Well, how fast is their clock ticking?" Or, "How old are they now?" Not allowed to say that according to relativity. 'Cause space and time is treated the same, and so it doesn't even make sense.

Sean Carroll
Yeah.

Lex Fridman
What happens to time in the holographic principle?

Sean Carroll
As far as we know, nothing dramatic happens. We're not anywhere close to being confident that we know what's going on here yet. So there are good unanswered questions about whether time is fundamental, whether time is emergent, whether it has something to do with quantum entanglement, whether time really exists at all, different theories, different proponents of different things, but there's nothing specifically about holography that would make us change our opinions about time, whatever they happen to be.

Lex Fridman
But holography is fundamentally about, it's a question of space?

Sean Carroll
It really is, yeah.

Lex Fridman
Okay. So time is just like an-

Sean Carroll
Time just goes along for the ride as far as we know. Yeah.

Lex Fridman
So all the questions about time is just almost like separate questions, whether it's emergent and all that kind of stuff?

Sean Carroll
Yeah, that might be a reflection of our ignorance right now, but yes.

Lex Fridman
If we figure out a lot, millions of years from now about black holes, how surprised would you be if they traveled back in time and told you everything you want to know about black holes? How much do you think there is still to know, and how mind-blowing would it be?

Sean Carroll
It does depend on what they would say. I think that there are colleagues of mine who think that we're pretty close to figuring out how information gets out of black holes, how to quantize gravity, things like that. I'm more skeptical that we are pretty close. I think that

there's room for a bunch of surprises to come. So in that sense, I suspect I would be surprised. The biggest and most interesting surprise to me would if quantum mechanics itself were somehow superseded by something better. As far as I know, there's no empirical evidence-based reason to think that quantum mechanics is not 100% correct, but it might not be. That's always possible, and there are, again, respectable friends of mine who speculate about it. So that's the first thing I'd want to know.

Lex Fridman
Oh, so the black hole would be the most clear illustration-

Sean Carroll
Yeah, that's where it would show up.

Lex Fridman
... or if there's something new it would show up there.

Sean Carroll
Maybe. The point is that black holes are mysterious for various reasons. So yeah, if our best theory of the universe is wrong, that might help explain why.

Lex Fridman
But do you think it's possible we'll find something interesting, like black holes sometimes create new universes or black holes are a kind of portal through space-time to another place or something like this. Then our whole conception of what is the fabric of space-time changes completely 'cause black holes, it's like Swiss cheese type of situation.

Sean Carroll
Yeah. That would be less surprising to me 'cause I've already written papers about that. We don't have, again, strong reason to think that the interior of a black hole leads to another universe. But it is possible, and it's also very possible that that's true for some black holes and not others. This is stuff, it's easy to ask questions we don't know the answer to. The problem is the questions that are easy to ask that we don't know the answer to are super hard to answer.

Lex Fridman
Because these objects are very difficult to test and to explore for us-

Sean Carroll
The regimes are just very far away. So either literally far away in space, but also in energy or mass or time or whatever.

Lex Fridman
You've published a paper on the holographic principle or that involves the holographic principle. Can you explain the details of that?

Sean Carroll
Yeah, I'm always interested in, since my first published paper, taking these wild speculative ideas and trying to test them against data. The problem is when you're dealing with wild speculative ideas, they're usually not well-defined enough to make a prediction. It's kind of, "I know what's going to happen in some cases, I don't know what's going to happen in other cases." So we did the following thing: As I've already mentioned, the holographic principle, which is meant to reflect the information contained in black holes seems to be telling us that there's less information, less stuff that can go on than you might naively expect. So let's upgrade naively expect to predict using quantum field theory. Quantum field theory is our best theory of fundamental physics right now. Unlike this holographic black hole stuff, quantum field theory is entirely local. In every point of space, something can go on. Then you add up all the different points in space, okay? Not holographic at all. So there's a mismatch between the expectation for what is happening even in empty space in quantum field theory versus what the holographic principle would predict. How do you reconcile these two things? So there's one way of doing it that had been suggested previously, which is to say that in the quantum field theory way of talking, it implies there's a whole bunch more states, a whole bunch more ways the system could be than there really are. I'll do a little bit of math just because there might be some people in the audience who like the math. If I draw two axes on a two-dimensional geometry, like the surface of the table, you know that the whole point of it being two-dimensional is I can draw two vectors that are perpendicular to each other. I can't draw three vectors that are all perpendicular to each other. They need to overlap a little bit. That's true for any numbers of dimensions. But I can ask, "Okay, how much do they have to overlap? If I try to put more vectors into a vector space, then the dimensionality of the vector space, can I make them almost perpendicular to each other?" The mathematical answer is, as the number of dimensions gets very, very large, you can fit a huge extra number of vectors in that are almost perpendicular to each other. So in this case, what we're suggesting is the number of things that can happen in a region of space is correctly described by holography. It is somewhat over-counted by quantum field theory, but that's because the quantum field theory states are not exactly perpendicular to each other. I should have mentioned that in quantum mechanics, states are given by vectors in some huge dimensional vector space; very, very, very, very large dimensional vector space. So maybe the quantum field theory states are not quite perpendicular to each other. If that is true, that's a speculation already. But if that's true, how would you know what is the experimental deviation? It would've been completely respectable if we had gone through and made some guesses and found that there is no noticeable experimental difference because, again, these things are in regimes very, very far away. We stuck our necks out. We made some very, very specific guesses as to how this weird overlap of states would show up in the equations of motion for particles like

neutrinos. Then we made predictions on how the neutrinos would behave on the basis of those wild guesses and then we compared them with data. What we found is we're pretty close but haven't yet reached the detectability of the effect that we are predicting. In other words, well, basically one way of saying what we predict is if a neutrino, and there's reasons why it's neutrinos, we can go into if you want, but it's not that interesting, if a neutrino comes to us from across the universe from some galaxy very, very far away, there is a probability as it's traveling that it will dissolve into other neutrinos because they're not really perpendicular to each other as vectors as they would ordinarily be in quantum field theory. That means that if you look at neutrinos coming from far enough away with high enough energies, they should disappear. If you see a whole bunch of nearby neutrinos, but then further away you should see fewer. There is an experiment called IceCube, which is this amazing testament to the ingenuity of human beings where they go to Antarctica and they drill holes and they put photodetectors on a string a mile deep in these holes. They basically use all of the ice in a cube, I don't know whether it's a mile or not, but it's like a kilometer or something like that, some big region. That much ice is their detector. They're looking for flashes when a cosmic ray or neutrino or whatever hits a water molecule in the ice [inaudible 00:51:47]

Lex Fridman
Make flashes in the ice.

Sean Carroll
Yes-

Lex Fridman
... they're looking for-

Sean Carroll
... they're looking for flashes in the ice.

Lex Fridman
What does the detector of that look like?

Sean Carroll
It's a bunch of strings, many, many, many strings with 360 degree photodetectors. You will-

Lex Fridman
That's really cool.

Sean Carroll
It's extremely cool. They've done amazing work, and they find neutrinos.

Lex Fridman
So they're looking for neutrinos.

Sean Carroll
Yeah. So the whole point is most cosmic rays are protons because why? Because protons exist, and they're massive enough that you can accelerate them to very high energies. So high-energy cosmic rays tend to be protons. They also tend to hit the Earth's atmosphere and decay into other particles. So neutrinos on the other hand, punch right through, at least usually, to a great extent, so not just Antarctica, but the whole earth. Occasionally, a neutrino will interact with a particle here on earth, and there's neutrinos is going through your body all the time from the sun, from the universe, etc. So if you're patient enough and you have a big enough part of the Antarctic ice sheet to look at, the nice thing about ice is it's transparent, so nature has built you a neutrino detector. That's what IceCube does.

Lex Fridman
So why ice? So is it just because the low noise and you get to watch this thing and it's-

Sean Carroll
It's much more dense than air, but it's transparent.

Lex Fridman
So yeah, much more dense, so higher probability, and then it's transparency, and then it's also in the middle of nowhere, so you can... Humans are great-

Sean Carroll
That's all you need. There's not that much ice-

Lex Fridman
I love it-

Sean Carroll
... right? Yeah.

Lex Fridman
... so humor me impressed.

Sean Carroll
There's more ice in Antarctic than anywhere else. Right. So anyway, you can go and you can get a plot from the IceCube experiment, how many neutrinos there are that they've detected with very high energies. We predict in our weird little holographic guessing game that there should be a cutoff. You should see neutrinos as you get to higher and higher energies and then they should disappear. If you look at the data, their data gives out exactly

where our cutoff is. That doesn't mean that our cutoff is right, it means they lose the ability to do the experiment exactly where we predict the cutoff should be.

Lex Fridman
Oh, boy, okay, but why is there a limit?

Sean Carroll
Oh, just because there are fewer, fewer high-energy neutrinos. So there's a spectrum and it goes down, but what we're plotting here is-

Lex Fridman
Got it.

Sean Carroll
... number of neutrinos versus energy, it's fading away, and they just get very, very few.

Lex Fridman
You need the high-energy neutrinos for your prediction.

Sean Carroll
Our effect is a little bit bigger for higher energies, yeah.

Lex Fridman
Got it, and that effect has to do with this almost perpendicular thing.

Sean Carroll
Let me just mention the name of Oliver Friedrich, who was a post-doc who led this. He deserves the credit for doing this. I was a co-author and a collaborator and I did some work, but he really gets the lion's share.

Lex Fridman
Thank you, Oliver. Thank you for pushing this wild science forward. Just to speak to that, the meta process of it, how do you approach asking these big questions and trying to formulate as a paper, as an experiment that could make a prediction, all that kind of stuff? What's your process?

Sean Carroll
There's very interesting things that happens once you're a theoretical physicist, once you become trained. You're a graduate student, you've written some papers and whatever, suddenly you are the world's expert in a really infinitesimally tiny area of knowledge and you know not that much about other areas. There's an overwhelming temptation to just drill deep, just keep doing basically the thing that you started doing, but maybe that thing you

started doing is not the most interesting thing to the world or to you or whatever. So you need to separately develop the capability of stepping back and going, " Okay, now that I can write papers in that area, now that I'm trained enough in the general procedure, what is the best match between my interests, my abilities and what is actually interesting?" Honestly, I've not been very good at that over my career. My process traditionally was I was working in this general area of particle physics, field theory, general relativity, cosmology, and I would try to take things other people were talking about and ask myself whether or not it really fit together. So I guess I have three papers that I've ever written that have done super well in terms of getting cited and things like that. One was my first ever paper that I get very little credit for, that was my advisor and his collaborator set that up. The other two were basically, my idea. One was right after we discovered that the universe was accelerating. So in 1998 observations showed that not only is the universe expanding, but it's expanding faster and faster. So that's attributed to either Einstein's cosmological constant or some more complicated form of dark energy, some mysterious thing that fills the universe. People were throwing around ideas about this dark energy stuff, "What could it be?" And so forth. Most of the people throwing around these ideas were cosmologists. They work on cosmology. They think about the universe all at once. Since I like to talk to people in different areas, I was more familiar than average with what a respectable working particle physicist would think about these things. What I immediately thought was, "You guys are throwing around these theories. These theories are wildly unnatural. They're super finely tuned. Any particle physicist would just be embarrassed to be talking about this." But rather than just scoffing at them, I sat down and asked myself, "Okay, is there a respectable version? Is there a way to keep the particle physicists happy but also make the universe accelerate?" I realized that there is some very specific set of models that is relatively natural, and guess what? You can make a new experimental prediction on the basis of those, and so I did that. People were very happy about that.

Lex Fridman
What was the thing that would make physicists happy that would make sense of this fragile thing that people call dark energy?

Sean Carroll
So the fact that dark energy pervades the whole universe and is slowly changing, that should immediately set off alarm bells because particle physics is a story of length scales and time scales that are generally, guess what? Small, right? Particles are small. They vibrate quickly, and you're telling me now I have a new field and its typical rate of change is once every billion years. That's just not natural. Indeed, you can formalize that and say, look, even if you wrote down a particle that evolved slowly over billions of years, if you let it interact with other particles at all, that would make it move faster, its dynamics would be faster, its mass would be higher, et cetera, et cetera. So there's a whole story. Things need to be robust, and they all talk to each other in quantum field theory. So how do you stop that from happening? The answer is symmetry. You can impose a symmetry that protects your

new field from talking to any other fields, and this is good for two reasons. Number one, it can keep the dynamics slow. So you can't tell me why it's slow. You just made that up, but at least it can protect it from speeding up because it's not talking to any other particles. The other is, it makes it harder to detect. Naively, experiments looking for fifth forces or time changes of fundamental constants of nature like the charge of the electron, these experiments should have been able to detect these dark energy fields, and I was able to propose a way to stop that from happening.

Lex Fridman
The detection.

Sean Carroll
The detection, yeah, because a symmetry could stop it from interacting with all these other fields, and therefore, it makes it harder to detect. Just by luck, I realized, 'cause it was actually based on my first-ever paper, there's one loophole. If you impose these symmetries, so you protect the dark energy field from interacting with any other fields, there's one interaction that is still allowed that you can't rule out. It is a very specific interaction between your dark energy field and photons, which are very common, and it has the following effect: As a photon travels through the dark energy, the photon has a polarization, up, down, left, right, whatever it happens to be, and as it travels through the dark energy, that photon will rotate its polarization. This is called birefringence. You can run the numbers and say you can't make a very precise prediction, 'cause we're making up this model. But if you want to roughly fit the data, you can predict how much polarization, rotation, there should be, a couple of degrees, not that much. So that's very hard to detect. People have been trying to do it. Right now, literally, we're on the edge of either being able to detect it or rule it out using the cosmic microwave background. There is just truth in advertising, there is a claim on the market that it's been detected, that it's there. It's not very statistically significant. If I were to bet, I think it would probably go away. It's very hard thing to observe. But maybe as you get better and better data, cleaner and cleaner analysis, it will persist, and we will have directly detected the dark energy.

Lex Fridman
So if we just take this tangent of dark energy, people will sometimes bring up dark energy and dark matter as an example why physicists have lost it, lost their mind. We're just going to say that there's this field that permeates everything. It's unlike any other field, and it's invisible, and it helps us work out some of the math. How do you respond to those kinds of suggestions.

Sean Carroll
Well, two ways. One way is, those people would've had to say the same thing when we discovered the planet Neptune, 'cause it's exactly analogous where we have a very good theory, in that case, Newtonian gravity in the solar system. We made predictions. The

predictions were slightly off for the motion of the outer planets. You found that you could explain that motion by positing something very simple, one more planet in a very, very particular place, and you went and looked for it, and there it was. That was the first successful example of finding dark matter in the universe.

Lex Fridman
It's a matter, though, we can't see.

Sean Carroll
Neptune was dark.

Lex Fridman
Yeah.

Sean Carroll
There's a difference between dark matter and dark energy. Dark matter as far as we are hypothesizing it is a particle of some sort. It's just a particle that interacts with us very weakly. So we know how much of it there is. We know more or less where it is. We know some of its properties. We don't know specifically what it is. But it's not anything fundamentally mysterious, it's a particle. Dark energy is a different story. So dark energy is indeed uniformly spread throughout space and has this very weird property that it doesn't seem to evolve as far as we can tell. It's the same amount of energy in every cubic centimeter of space from moment to moment in time. That's why far and away the leading candidate for dark energy is Einstein's cosmological constant. The cosmological constant is strictly constant, 100% constant. The data say it better be 98% constant or better, so 100% constant works, and it's also very robust. It's just there. It's not doing anything. It doesn't interact with any other particles. It makes perfect sense. Probably the dark energy is the cosmological constant. The dark matter, super important to emphasize here. It was hypothesized at first in the '70s and '80s mostly to explain the rotation of galaxies. Today, the evidence for dark matter is both much better than it was in the 1980s and from different sources. It is mostly from observations of the cosmic background radiation or of large scale structure. From observations of the cosmic background radiation or of large-scale structure. We have multiple independent lines of evidence, also gravitational lensing and things like that, many, many pieces of evidence that say that dark matter is there and also that say that the effects of dark matter are different than if we modified gravity. That was my first answer to your question is dark matter we have a lot of evidence for. But the other one is of course we would love it if it weren't dark matter. Our vested interest is 100% aligned with it being something more cool and interesting than dark matter because dark matter's just a particle. That's the most boring thing in the world.

Lex Fridman
And it's non-uniformly distributed through space, dark matter?

Sean Carroll
Absolutely. Yeah.

Lex Fridman
And so this-

Sean Carroll
You can even see maps of it that we've constructed from gravitational lensing.

Lex Fridman
Verifiable clumps of dark matter in the galaxy that explains stuff.

Sean Carroll
Bigger than the galaxy, sadly. We think that in the galaxy dark matter is lumpy, but it's weaker, its effects are weaker. But on the scale of large scale structure and clusters of galaxies and things like that, yes, we can show you where the dark matter is.

Lex Fridman
Could there be a super cool explanation for dark matter that would be interesting as opposed to just another particle that sits there and clumps?

Sean Carroll
The super cool explanation would be modifying gravity rather than inventing a new particle. Sadly, that doesn't really work. We've tried. I've tried. That's my third paper that was very successful. I tried to unify dark matter and dark energy together. That was my idea. That was my aspiration, not even idea. I tried to do it. It failed even before we wrote the paper. I realized that my idea did not help. It could possibly explain away the dark energy, but it would not explain the way the dark matter, and so I thought it was not that interesting, actually. And then two different collaborators of mine said, "Has anyone thought of this idea?" They thought of exactly the same idea completely independently of me. And I said, "Well, if three different people found the same idea, maybe it is interesting," and so we wrote the paper. And yeah, it was very interesting. People are very interested in it.

Lex Fridman
Can you describe this paper a little bit? It's fascinating how much of a thing there is, dark energy and dark matter, and we don't quite understand it. What was your dive into exploring how to unify the two?

Sean Carroll
Here is what we know about dark matter and dark energy: They become important in regimes where gravity is very, very, very weak. That's the opposite from what you would expect if you actually were modifying gravity. There's a rule of thumb in quantum field

theory, et cetera that new effects show up when the effects are strong. We understand weak fields, we don't understand strong fields. But okay, maybe this is different. What do I mean by when gravity is weak? The dark energy shows up late of the universe. Early in the history of the universe, the dark energy is irrelevant, but remember the density of dark energy stays constant. The density of matter and radiation go down. At early times, the dark energy was completely irrelevant compared to matter and radiation. At late times, it becomes important. That's also when the universe is dilute and gravity is relatively weak. Now think about galaxies. A galaxy is more dense in the middle, less dense on the outside. And there is a phenomenological fact about galaxies that in the interior of galaxies you don't need dark matter. That's not so surprising because the density of stars and gas is very high there and the dark matter is just subdominant. But then there's generally a radius inside of which you don't need dark matter to fit the data, outside of which you do need dark matter to fit the data. That's again when gravity is weak. I asked myself, "Of course, we know in field theory new effects should show up when fields are strong, not weak, but let's throw that out of the window. Can I write down a theory where gravity alters when it is weak?" And we've already said what gravity is. What is gravity? It's the curvature of space-time. There are mathematical quantities that measure the curvature of space-time. And generally, you would say, "I have an understanding, Einstein's equation," which I explained to the readers in the book, "relates the curvature of space-time to matter and energy. The more matter and energy, the more curvature." I'm saying what if you add a new term in there that says, "The less matter and energy, the more curvature"? No reason to do that except to fit the data. I tried to unify the need for dark matter and the need for dark energy.

Lex Fridman
That would be really cool if that was the case.

Sean Carroll
Super cool. It'd be the best. It'd be great. It didn't work.

Lex Fridman
It'd be really interesting if gravity did something funky when there's not much of it, almost like at the edges of it gets noisy.

Sean Carroll
That was exactly the hope.

Lex Fridman
Right. Aw, man.

Sean Carroll
But the great thing about physics is there are equations. You can come up with the words and you can wave your hands, but then you got to write down the equations; and I did. And I

figured out that it could help with the dark energy, the acceleration of the universe; it doesn't help with dark matter at all. Yeah.

Lex Fridman
It just sucks that the scale of galaxies and scale of solar systems, the physics is boring.

Sean Carroll
Yeah, it does. I agree. I tear my hair out when people who are not physicists accuse physicists, like you say, of losing the plot because they need dark matter and dark energy. I don't want dark matter and dark energy; I want something much cooler than that. I've tried. But you got to listen to the equations and to the data.

Lex Fridman
You've mentioned three papers, your first ever, your first awesome paper ever, and your second awesome paper ever. Of course you wrote many papers, so you're being very harsh on the others. But-

Sean Carroll
Well, by the way, this is not awesomeness, this is impact.

Lex Fridman
Impact.

Sean Carroll
Right?

Lex Fridman
Sure.

Sean Carroll
There's no correlation between awesomeness and impact. Some of my best papers fell without a stone and vice versa.

Lex Fridman
Tree falls in the forest. Yeah.

Sean Carroll
Yeah. The first paper was called Limits on the Lorentz and Parity Violating Modification of Electromagnetism... Or Electrodynamics. We figured out how to violate Lorentz invariance, which is the symmetry underlying relativity. And the important thing is we figured out a way to do it that didn't violate anything else and was experimentally testable. People love that. The second paper was called Quintessence and the Rest of the World. Quintessence is this

dynamical dark energy field. The rest of the world is because I was talking about how the quintessence field would interact with other particles and fields and how to avoid the interactions you don't want. And the third paper was called Is Cosmic Speed-Up Due to Gravitational Physics? Something like that. You see the common theme. I'm taking what we know, the standard model of particle physics, general relativity, tweaking them in some way, and then trying to fit the data

Lex Fridman
And trying to make it so it's experimentally validated.

Sean Carroll
Ideally, yes, that's right. That's the goal.

Lex Fridman
You wrote the book Something Deeply Hidden on the mysteries of quantum mechanics and a new book coming out soon, part of that, Biggest Ideas in the Universe series we mentioned called Quanta and Fields. That's focusing on quantum mechanics. Big question first, biggest ideas in the universe, what to you is most beautiful or perhaps most mysterious about quantum mechanics?

Sean Carroll
Quantum mechanics is a harder one. I wrote a textbook on general relativity, and I started it by saying, "General relativity is the most beautiful physical theory ever invented." And I will stand by that. It is less fundamental than quantum mechanics, but quantum mechanics is a little more mysterious. It's a little bit kludgy right now. If you think about how we teach quantum mechanics to our students, the Copenhagen interpretation, it's a God-awful mess. No one's going to accuse that of being very beautiful. I'm a fan of the many-worlds interpretation of quantum mechanics, and that is very beautiful in the sense that fewer ingredients, just one equation, and it could cover everything in the world. It depends on what you mean by beauty, but I think that the answer to your question is quantum mechanics can start with extraordinarily austere, tiny ingredients and in principle lead to the world. That boggles my mind. It's much more comprehensive. General relativity is about gravity, and that's great. Quantum mechanics is about everything and seems to be up to the task. And so I don't know, is that beauty or not? But it's certainly impressive.

Lex Fridman
Both for the theory, the predictive power of the theory and the fact that the theory describes tiny things creating everything we see around us.

Sean Carroll
It's a monist theory. In classical mechanics, I have a particle here, particle there; I describe them separately. I can tell you what this particle's doing, what that particle's doing. In

quantum mechanics, we have entanglement, as Einstein pointed out to us in 1935. And what that means is there is a single state for these two particles. There's not one state for this particle, one state for the other particle. And indeed, there's a single state for the whole universe called the wave function of the universe, if you want to call it that. And it obeys one equation. And is our job then to chop it up, to carve it up, to figure out how to get tables and chairs and things like that out of it.

Lex Fridman
You mentioned the many-worlds interpretation, and it is in fact beautiful, but it's one of your more controversial things you stand behind. You've probably gotten a bunch of flak for it.

Sean Carroll
I'm a big boy. I can take it.

Lex Fridman
Well, can you first explain it and then maybe speak to the flak you may have gotten?

Sean Carroll
Sure. The classic experiment to explain quantum mechanics to people is called the Stern-Gerlach experiment. You're measuring the spin of a particle. And in quantum mechanics, the spin is just a spin. It's the rate at which something is rotating around in a very down to earth sense, the difference being is that it's quantized. For something like a single electron or a single neutron, it's either spinning clockwise or counterclockwise. Let's put it this way. Those are the only two measurement outcomes you will ever get. There's no it's spinning faster or slower, it's either spinning one direction or the other. That's it. Two choices. According to the rules of quantum mechanics, I can set up an electron, let's say, in a state where it is neither purely clockwise or counterclockwise but a superposition of both. And that's not just because we don't know the answer, it's because it truly is both until we measure it. And then when we measure it, we see one or the other. This is the fundamental mystery of quantum mechanics is that how we describe the system when we're not looking at it is different from what we see when we look at it. We teach our students in the Copenhagen way of thinking is that the act of measuring the spin of the electron causes a radical change in the physical state. It spontaneously collapses from being a superposition of clockwise and counterclockwise to being one or the other. And you can tell me the probability that that happens, but that's all you can tell me. And I can't be very specific about when it happens, what caused it to happen, why it's happening, none of that. That's all called the measurement problem of quantum mechanics. Many-worlds just says, "Look, I just told you a minute ago that there's only one way function for the whole universe, and that means that you can't take too seriously just describing the electron, you have to include everything else in the universe." In particular, you clearly have to interact with the electron in order to measure it. Whatever is interacting with the electron should be included in the wave function that you're describing. And look, maybe it's just you, maybe

your eyeballs are able to perceive it, but okay, I'm going to include you in the wave function. Since you have a very sophisticated listenership, I'll be a little bit more careful than average. What does it mean to measure the spin of the electron? We don't need to go into details, but we want the following thing to be true: If the electron were in a state that was 100% spinning clockwise, then we want the measurement to tell us it was spinning clockwise. We want your brain to go, "Yes, the electron was spinning clockwise." Likewise, if it was 100% counterclockwise, we want to see that, to measure that. The rules of quantum mechanics, the Schrodinger equation of quantum mechanics, is 100% clear that if you want to measure it clockwise when it's clockwise and measure it counterclockwise when it's counterclockwise, then when it starts out in a superposition, what will happen is that you and the electron will entangle with each other. And by that I mean that the state of the universe evolves into part saying, "The electron was spinning clockwise, and I saw it clockwise," and part of the state is it's in a superposition with the part that says, "The electron was spinning counterclockwise, and I saw it counterclockwise." Everyone agrees with this; entirely uncontroversial. Straightforward consequence of the Schrodinger equation. And then Niels Bohr would say, "And then part of that wave function disappears," and we're in the other part. And you can't predict which part it'll be, only the probability. Hugh Everett, who was a graduate student in the 1950s, was thinking about this, says, "I have a better idea. Part of the wave function does not magically disappear, it stays there." The reason why that idea, Everett's idea that the whole wave function always sticks around and just obeys the Schrodinger equation was not thought of years before is because naively, you look at it and you go, "Okay, this is predicting that I will be in a superposition, that I will be in a superposition of having seen the electron be clockwise and having seen it be counterclockwise." No experimenter has ever felt like they were in a superposition. You always see an outcome. Everett's move, which was genius, was to say, "The problem is not the Schrodinger equation. The problem is you have misidentified yourself in the Schrodinger equation." You have said, "Oh, look, there's a person who saw counterclockwise, there's a person who saw clockwise; I should be that superposition of both." And Everett says, "No, no, no, you're not," because the part of the wave function in which the spin was clockwise, once that exists, it is completely unaffected by the part of the wave function that says the spin was counterclockwise. They are apart from each other. They are un-interacting. They have no influence. What happens in one part has no influence in the other part. Everett says, "The simple resolution is to identify yourself as either the one who saw spin clockwise or the one who saw spin counterclockwise." There are now two people once you've done that experiment. The Schrodinger equation doesn't have to be messed with, all you have to do is locate yourself correctly in the wave function. That's many-worlds.

Lex Fridman
The number of worlds is-

Sean Carroll
Very big.

Lex Fridman
... very, very, very big. Where do those worlds fit? Where do they go?

Sean Carroll
The short answer is the worlds don't exist in space, space exists separately in each world. There's a technical answer to your question, which is Hilbert space, the space of all possible quantum mechanical states, but physically, we want to put these worlds somewhere. That's just a wrong intuition that we have. There is no such thing as the physical spatial location of the worlds because space is inside the worlds.

Lex Fridman
One of the properties of this interpretation is that you can't travel from one world to the other.

Sean Carroll
That's right.

Lex Fridman
Which makes you feel that they're existing separately.

Sean Carroll
They are existing separately and simultaneously.

Lex Fridman
And simultaneously.

Sean Carroll
Without locations in space.

Lex Fridman
Without locations in space. How is it possible to visualize them existing without a location in space?

Sean Carroll
The real answer to that, the honest answer is the equations predict it. If you can't visualize it, so much worse for you. The equations are crystal clear about what they're predicting.

Lex Fridman
Is there a way to get closer to understanding and visualizing the weirdness of the implications of this?

Sean Carroll
I don't think it's that hard. It wasn't that hard for me. I don't mind the idea that when I make a quantum mechanical measurement there is, later on in the universe, multiple descendants of my present self who got different answers for that measurement. I can't interact with them. Hilbert space, the space law of quantum wave functions, was always big enough to include all of them. I'm going to worry about the parts of the universe I can observe. Let's put it this way. Many-worlds comes about by taking the Schrodinger equation seriously. The Schrodinger equation was invented to fit the data, to fit the spectrum of different atoms and different emission and absorption experiments. And it's perfectly legitimate to say, "Well, okay, you're taking the Schrodinger equation, you're extrapolating it, you're trusting it, believing it beyond what we can observe. I don't want to do that." That's perfectly legit except, okay, then what do you believe? Come up with a better theory. You're saying you don't believe the Schrodinger equation; tell me the equation that you believe in. And people have done that. Turns out it's super hard to do that in a legitimate way that fits the data.

Lex Fridman
And many-worlds is a really clean.

Sean Carroll
Absolutely the most austere, clean, no extra baggage theory of quantum mechanics.

Lex Fridman
But if it in fact is correct, isn't this the weirdest thing of anything we know?

Sean Carroll
Yes. In fact, let me put it this way. The single best reason in my mind to be skeptical about many-worlds is not because it doesn't make sense or it doesn't fit the data or I don't know where the worlds are going or whatever, it's because to make that extrapolation, to take seriously the equation that we know is correct in other regimes requires new philosophy, requires a new way of thinking about identity, about probability, about prediction, a whole bunch of things. It's work to do that philosophy, and I've been doing it and others have done it, and I think it's very, very doable, but it's not straightforward. It's not a simple extrapolation from what we already know, it's a grand extrapolation very far away. And if you just wanted to be methodologically conservative and say, "That's a step too far; I don't want to buy it," I'm sympathetic to that. I think that you're just wimping out, I think that you should have more courage, but I get the impulse.

Lex Fridman
And there is, under many-worlds, an era of time where, if you rewind it back, there's going to be one initial state.

Sean Carroll
That's right. All of quantum mechanics, all different versions require a kind of arrow of time. It might be different in every kind, but the quantum measurement process is irreversible. You can measure something, it collapses; you can't go backwards. If someone tells you the outcome... If I say I've measured an electron, "Its spin is clockwise," and they say, "What was it before I measured it?" You know there was some part of it that was clockwise, but you don't know how much. And many-worlds is no different. But the nice thing is that the kind of arrow of time you need in many-worlds is exactly the kind of arrow of time you need anyway for entropy and thermodynamics and so forth. You need a simple, low entropy initial state. That's what you need in both cases.

Lex Fridman
If you actually look at under many-worlds into the entire history of the universe, correct me if I'm wrong, but it looks very deterministic.

Sean Carroll
Yes.

Lex Fridman
In each moment, does the moment contain the memory of the entire history of the universe? To you, does the moment contain the memory of everything that preceded it?

Sean Carroll
As far as we know, according to many-worlds, the wave function of the universe, all the branches of the universe at once, all the worlds does contain all the information. Calling it a memory is a little bit dangerous because it's not the same kind of memory that you and I have in our brains because our memories rely on the arrow of time, and the whole point of the Schrodinger equation or Newton's laws is they don't have an arrow of time built in. They're reversible. The state of the universe not only remembers where it came from but also determines where it's going to go in a way that our memories don't do that.

Lex Fridman
But our memories, we can do replay. Can you do this?

Sean Carroll
We can, but the act of forming a memory increases the entropy of the universe. It is an irreversible process also. You can walk on a beach and leave your footprints there. That's a

record of your passing. It will eventually be erased by the ever-increasing entropy of the universe.

Lex Fridman
Well, but you can imperfectly replay it. I guess can we return, travel back in time imperfectly?

Sean Carroll
Oh, it depends on the level of precision you're trying to ask that question. The universe contains the information about where the universe was, but you and I don't. We're nowhere close.

Lex Fridman
And it's, what, computationally very costly to try to consult the universe?

Sean Carroll
Well, it depends on, again, exactly what you're asking. There are some simple questions like what was the temperature of the universe 30 seconds after the Big Bang? We can answer that. That's amazing that we can answer that to pretty high precision. But if you want to know where every atom was, then no.

Lex Fridman
What to you is the Big Bang? Why? Why did it happen?

Sean Carroll
We have no idea. I think that that's a super important question that I can imagine making progress on, but right now I'm more or less maximally uncertain about what the answer is.

Lex Fridman
Do you think black holes will help potentially?

Sean Carroll
No.

Lex Fridman
No.

Sean Carroll
Not that much. Quantum gravity will help, and maybe black holes will help us figure out quantum gravity, so indirectly, yes. But we have the situation where general relativity, Einstein's theory unambiguously predicts there was a singularity in the past. There was a moment of time when the universe had infinite curvature, infinite energy, infinite expansion

rate, the whole bit. That's just a fancy way of saying the theory has broken down. And classical general relativity is not up to the task of what saying what really happened at that moment. It is completely possible there was, in some sense, a moment of time before which there were no other moments. And that would be the Big Bang. Even if it's not a classical general relativity kind of thing, even if quantum mechanics is involved, maybe that's what happened. It's also completely possible there was time before that space and time and they evolved into our hot big bang by some procedure that we don't really understand.

Lex Fridman
And if time and space are emergent, then the before even starts getting real weird.

Sean Carroll
Well, I think that if there is a first moment of time, that would be very good evidence or that would fit hand in glove with the idea that time is emergent. If time is fundamental, then it tends to go forever because it's fundamental.

Lex Fridman
Well, yeah. The general formulation of this question is what's outside of it? Well, what's outside of our universe, in time and in space? I know it's a pothead question, Sean. I understand. I apologize.

Sean Carroll
That's my life. My life is asking pothead questions. Some of them, the answer is that's not the right way to think about it.

Lex Fridman
Okay. But is it possible to think at all about what's outside our universe?

Sean Carroll
It's absolutely legit to ask questions, but you have to be comfortable with the possibility that the answer is there's no such thing as outside our universe. That's absolutely on the table. In fact, that is the simplest, most likely to be correct answer that we know of.

Lex Fridman
But it's the only thing in the universe that wouldn't have an outside.

Sean Carroll
Yeah. If the universe is the totality of everything, it would not have an outside.

Lex Fridman
That's so weird to think that there's not an outside. We want there to be a creator, a creative force that led to this and an outside. This is our town, and then there's a bigger world. And there's always a bigger world. And to think that there's not [inaudible 01:29:53].

Sean Carroll
Because that is our experience. That's the world we grew up in. The universe doesn't need to obey those rules.

Lex Fridman
Such a weird thing.

Sean Carroll
When I was a kid, that used to keep me up at night. What if the universe had not existed?

Lex Fridman
Right. It feels like a lot of pressure that if this is the only universe and we're here, one of the few intelligent civilizations, maybe the only one, it's the old theories that we're the center of everything, it just feels suspicious. That's why many-worlds is exciting to me because it is humbling in all the right kinds of ways. It feels like infinity is the way this whole thing runs.

Sean Carroll
There's one pitfall that I'll just mention because there's a move that is made in these theoretical edges of cosmology that I think is a little bit mistaken, which is to say I'm going to think about the universe on the basis of imagining that I am a typical observer. This is called the principle of typicality, or the principle of mediocrity, or even the Copernican principle. Nothing special about me, I'm just typical in the universe. But then you draw some conclusions from this, and what you end up realizing is you've been hilariously presumptuous because by saying, "I'm a typical observer in the universe," you're saying, "Typical observers in the universe are like me," and that is completely unjustified by anything. I'm not telling you what the right way to do it is, but these kinds of questions that are not quite grounded in experimental verification or falsification are ones you have to be very careful about.

Lex Fridman
That to me is one of the most interesting questions. And there's different ways to approach it, but what's outside of this? How did the big mess start? How do we get something from nothing? That's always the thing you're sneaking up to when you're studying all of these questions. You're always thinking that's where the black hole and the unifying, getting quantum gravity, all this kind of stuff, you're always sneaking up to that question, where did all of this come from?

Sean Carroll
Yeah, that's fair.

Lex Fridman
And I think that's probably an answerable question, right?

Sean Carroll
No.

Lex Fridman
It doesn't have to be. You think there could be a turtle at the bottom of this that refuses to reveal its identity?

Sean Carroll
Yes. I think that specifically the question why is there something rather than nothing? does not have the kind of answer that we would ordinarily attribute to why questions because typical why questions are embedded in the universe. And when we answer them, we take advantage of the features of the universe that we know and love. But the universe itself, as far as we know, is not embedded in anything bigger or stronger, and therefore it can just be.

Lex Fridman
Do you think it's possible this whole place is simulated?

Sean Carroll
Sure.

Lex Fridman
It's a really interesting, dark, twisted video game that we're all existing in.

Sean Carroll
My own podcast listeners, Mindscape listeners tease me because they know from my AMA episodes that if you ever start a question by asking, "Do you think it's possible that..." the answer's going to be yes. That might not be the answer that you care about, but it's possible, sure, as long as you're not adding two even numbers together and getting an odd number.

Lex Fridman
When you say it's possible, there's a mathematically yes, and then there's more of intuitive.

Sean Carroll
Yeah. You want to know whether it's plausible. You want to know is there a-

Lex Fridman
Plausible.

Sean Carroll
... reasonable, non-zero credence to attach to this? I don't think that there's any philosophical knockout objection to the simulation hypothesis. I also think that there's absolutely no reason to take it seriously.

Lex Fridman
Do you think humans will try to create one? I guess that's how I always think about it. I've spent quite a bit of time over the past few years and a lot more recently in virtual worlds and just am always captivated by the possibility of creating higher and high resolution worlds. And as we'll talk a little bit about artificial intelligence, the advancement on the Sora front, you can automatically generate those worlds, and the possibility of existing in those automatically generated worlds is pretty exciting as long as there's a consistent physics, quantum mechanics and general relativity that governs the generation of those worlds. It just seems like humans will for sure try to create this.

Sean Carroll
Yeah, I think they will create better and better simulations. I think the philosopher, David Chalmers, has done what I consider to be a good job of arguing that we should treat things that happen in virtual reality and in simulated realities as just as real as the reality that we experience. I also think that as a practical matter, people will realize how much harder it is to simulate a realistic world than we naively believe. This is not a my lifetime kind of worry.

Lex Fridman
Yeah. The practical matter of going from a prototype that's impressive to a thing that governs everything. Similar question on this front is in AGI. You've said that we're very far away from AGI.

Sean Carroll
I want to eliminate the phrase AGI.

Lex Fridman
Basically, when you're analyzing large language models and seeing how far are they from whatever AGI is, and we can talk about different notions of intelligence, that we're not as close as some people in public view are talking about. What's your intuition behind that?

Sean Carroll
My intuition is basically that artificial intelligence is different than human intelligence, and so the mistake that is being made by focusing on AGI among those who do is an artificial agent that, as we can make them now or in the near future, might be way better than human

beings at some things, way worse- ... Better than human beings at some things. Way worse than human beings at other things. And rather than trying to ask, how close is it to being a human-like intelligent, we should appreciate it for what its capabilities are, and that will both be more accurate and help us put it to work and protect us from the dangers better rather than always anthropomorphizing it.

Lex Fridman
I think the underlying idea there under the definition of AGI is that the capabilities are extremely impressive. That's not a precise statement, but meaning-

Sean Carroll
Sure. No, I get that. I completely agree.

Lex Fridman
And then the underlying question where a lot of the debate is, is how impressive is it? What are the limits of large language models? Can they really do things like common sense reasoning? How much do they really understand about the world or are they just fancy mimicry machines? And where do you fall on that as to the limits of large language models?

Sean Carroll
I don't think that there are many limits in principle. I am a physicalist about consciousness and awareness and things like that. I see no obstacle to, in principle, building an artificial machine that is indistinguishable in thought and cognition from a human being. But we're not trying to do that. What a large language model is trying to do is to predict text. That's what it does. And it is leveraging the fact that we human beings for very good evolutionary biology reasons, attribute intentionality and intelligence and agency to things that act like human beings. As I was driving here to get to this podcast space, I was using Google Maps and Google Maps was talking to me, but I wanted to stop to get a cup of coffee. So I didn't do what Google Maps told me to do. I went around a block that it didn't like. And so it gets annoyed. It says like, "No, why are you doing ..." It doesn't say exactly in this, but you know what I mean. It's like, "No, turn left, turn left," and you turn right. It is impossible as a human being not to feel a little bit sad that Google Maps is getting mad at you. It's not. It's not even trying to, it's not a large language model, no aspirations to intentionality, but we attribute that all the time. Dan Dennett, the philosopher, wrote a very influential paper on The Intentional Stance, the fact that it's the most natural thing in the world for we human beings to attribute more intentionality to artificial things than are really there, which is not to say it can't be really there. But if you're trying to be rational and clear thinking about this, the first step is to recognize our huge bias towards attributing things below the surface to systems that are able to, at the surface level, act human.

Lex Fridman
So if that huge bias of intentionality is there in the data, in the human data, in the vast landscape of human data that AI models, large language models, and video models in the future are trained on, don't you think that that intentionality will emerge as fundamental to the behavior of these systems naturally?

Sean Carroll
Well, I don't think it will happen naturally. I think it could happen. Again, I'm not against the principle. But again, the way that large language models came to be and what they're optimized for is wildly different than the way that human beings came to be and what they're optimized for. So I think we're missing a chance to be much more clear-headed about what large language models are by judging them against human beings. Again, both in positive ways and negative ways.

Lex Fridman
Well, I think ... To push back on what they're optimized for is different to describe how they're trained versus what they're optimized for. So they're trained in this very trivial way of predicting text tokens, but you can describe what they're optimized for and what the actual task in hand is, is to construct a world model, meaning an understanding of the world. And that's where it starts getting closer to what humans are kind of doing, where just in the case of large language models, know how the sausage is made, and we don't know how it's made for us humans.

Sean Carroll
But they're not optimized for that. They're optimized to sound human.

Lex Fridman
That's the fine-tuning. But the actual training is optimized for understanding, creating a compressed representation of all the stuff that humans have created on the internet.

Sean Carroll
Right.

Lex Fridman
And the hope is that that gives you a deep understanding of the world.

Sean Carroll
Yeah. So that's why I think that there's a set of hugely interesting questions to be asked about the ways in which large language models actually do represent the world. Because what is clear is that they're very good at acting human. The open question in my mind is, is the easiest, most efficient, best way to act human to do the same things that human beings do or are there other ways? And I think that's an open question. I just heard a talk by Melanie

Mitchell at Santa Fe Institute, an artificial intelligence researcher, and she told two stories about two different papers, one that someone else wrote and one that her group is following up on. And they were modeling Othello. Othello, the game with a little rectangular board, white and black squares. So the experiment was the following. They fed a neural network the moves that were being made in the most symbolic form, E5 just means that, okay, you put a token down on E5. So it gives a long string, it does this for millions of games, real legitimate games. And then it asks the question, the paper asks the question, "Okay, you've trained it to tell what would be a legitimate next move from not a legitimate next move. Did it in its brain, in its little large language model brain." I don't even know if it's technically large language model, but a deep learning network. "Did it come up with a representation of the Othello board?" Well, how do you know? And so they construct a little probe network that they insert, and you ask it, "What is it doing right at this moment?" And the answer is that the little probe network can ask, "Would this be legitimate or is this token white or black?" Or whatever, things that in practice would amount to it has invented the Othello board. And it found that the probe got the right answer, not 100% of the time, but more than by chance, substantially more than by chance. So they said there's some tentative evidence that this neural network has discovered the Othello board just out of data, raw data. But then Melanie's group asked the question, "Okay, are you sure that that understanding of the Othello board wasn't built into your probe?" And what they found was at least half of the improvement was built into the probe. Not all of it. And look, a Othello board is way simpler than the world. So that's why I just think it's an open question, whether or not ... I mean, it would be remarkable either way to learn that large language models that are good at doing what we train them to do are good because they've built the same kind of model of the world that we have in our minds or that they're good despite not having that model. Either one of these is an amazing thing. I just don't think the data are clear on which one is true.

Lex Fridman
I think I have some sort of intellectual humility about the whole thing because I was humbled by several stages in the machine learning development over the past 20 years. And I just would never have predicted that LLMs, the way they're trained, on the scale of data they're trained would be as impressive as they are. And that's where intellectual humility steps in, where my intuition would say something like with Melanie, where you need to be able to have very sort of concrete common sense reasoning, symbolic reasoning type things in a system in order for it to be very intelligent. But here, I'm so impressed by what it's capable to do, train on the next token prediction essentially ... My conception of the nature of intelligence is just completely, not completely, but humbled, I should say.

Sean Carroll
Look, and I think that's perfectly fair. I also was, I almost say pleasantly, but I don't know whether it's pleasantly or unpleasantly, but factually surprised by the recent rate of progress. Clearly some kind of phase transition percolation has happened and the

improvement has been remarkable, absolutely amazing. That I have no arguments with. That doesn't yet tell me the mechanism by which that improvement happened. Constructing a model much like a human being is clearly one possible mechanism, but part of the intellectual humility is to say maybe there are others.

Lex Fridman
I was chatting with the CEO of Anthropic, Dario Amodei, so behind Claude and that company, but a lot of the AI companies are really focused on expanding the scale of compute. If we assume that AI is not data limited, but is compute limited, you can make the system much more intelligent by using more compute. So let me ask you almost on the physics level, do you think physics can help expand the scale of compute and maybe the scale of energy required to make that compute happen?

Sean Carroll
Yeah, 100%. I think this is one of the biggest things that physics can help with, and it's an obvious kind of low-hanging fruit situation where the heat generation, the inefficiency, the waste of existing high-level computers is nowhere near the efficiency of our brains. It's hilariously worse, and we haven't tried to optimize that hard on that frontier. I mean, your laptop heats up when it's sitting on your lap. It doesn't need to. Your brain doesn't heat up like that. So clearly there exists in the world of physics, the capability of doing these computations with much less waste heat being generated, and I look forward to people doing that, yeah.

Lex Fridman
Are you excited for the possibility of nuclear fusion?

Sean Carroll
I am cautiously optimistic. Excited would be too strong. I mean, it'd be great, but if we really tried solar power, it would also be great.

Lex Fridman
I think Ilya Sutskever said this, that the future of humanity on Earth will be just the entire surface of Earth is covered in solar panels and data centers.

Sean Carroll
Why would you waste the surface of the Earth with solar panels? Put them in space.

Lex Fridman
Sure, you can go in space. Yeah.

Sean Carroll
Space is bigger than the Earth.

Lex Fridman
Yeah, just solar panels everywhere.

Sean Carroll
Yeah.

Lex Fridman
I like it.

Sean Carroll
We already have fusion. It's called the Sun.

Lex Fridman
Yeah, that's true. And there's probably more and more efficient ways of catching that energy.

Sean Carroll
Sending it down is the hard part, absolutely. But that's an engineering problem.

Lex Fridman
So I just wonder where the data centers, the compute centers can expand to, if that's the future. If AI is as effective as it possibly could be, then the scale of computation will keep increasing, but perhaps it's a race between efficiency and scale.

Sean Carroll
There are constraints. There's a certain amount of energy, a certain amount of damage we can do to the environment before it's not worth it anymore. So yeah, I think that's a new question. In fact, it's kind of frustrating because we get better and better at doing things efficiently, but we invent more things we want to do faster than we get good at doing them efficiently. So we're continuing to make things worse in various ways.

Lex Fridman
I mean, that's the dance of humanity where we're constantly creating better motivated technologies that are potentially causing a lot more harm, and that includes for weapons, includes AI used as weapons, that includes nuclear weapons, of course, which is surprising to me that we haven't destroyed human civilization yet, given how many nuclear warheads are out there.

Sean Carroll
Look, I'm with you. Between nuclear and bioweapons, it is a little bit surprising that we haven't caused enormous devastation. Of course, we did drop two atomic bombs on Japan,

but compared to what could have happened or could happen tomorrow, it could be much worse.

Lex Fridman
It does seem like there's an underlying, speaking of quantum fields, there's a field of goodness within the human heart that in some kind of game theoretic way, we create really powerful things that could destroy each other, and there's greed and ego and all this kind of power hungry dictators that are at play here in all the geopolitical landscape, but we somehow always don't go too far.

Sean Carroll
But that's exactly what you would say right before we went too far.

Lex Fridman
Right before we went too far, and that's why we don't see aliens. So you're like I mentioned, associated with Santa Fe Institute. I just would love to take a stroll down the landscape of ideas explored there.

Sean Carroll
Sure.

Lex Fridman
So they look at complexity in all kinds of ways. What do you think about the emergence of complexity from simple things interacting simply?

Sean Carroll
I think it's a fascinating topic. I mean, that's why I'm thinking about these things these days rather than the papers that I was describing to you before. All of those papers I described to you before are guesses. What if the laws of physics are different in the following way? And then you can work out the consequences. At some point in my life, I said, "What is the chance I'm going to guess right?" Einstein guessed right, Steven Weinberg guessed right, but there's a very small number of times that people guessed right. Whereas with this emergence of complexity from simplicity, I really do think that we haven't understood the basics yet. I think we're still kind of pre-paradigmatic. There have been some spectacular discoveries. People like Geoffrey West at Santa Fe and others have really given us true insights into important systems. But still, there's a lot of the basics, I think are not understood. And so searching for the general principles is what I like to do, and I think it's absolutely possible that ... And to be a little bit more substantive than that. This is kind of a cliche. I think the key is information, and I think that what we see through the history of the universe as you go from simple to more and more complex is really subsystems of the universe figuring out how to use information to do whatever, to survive or to thrive or to

reproduce. I mean, that's the sort of fuel, the leverage, the resource that we have for a while anyway, until the heat death. But that's where the complexity is really driven by.

Lex Fridman
But the mechanism of it. I mean, you mentioned Geoffrey West. What are interesting inklings of progress in this realm? And what are systems that interest you in terms of information? So I mean, for me, just as a fan of complexity, just even looking at simple cellular automata is always just a fascinating way to illustrate the emergence of complexity.

Sean Carroll
So for those of the listeners who don't know, viewers, cellular automata come from imagining a very simple configuration. For example, a set of ones and zeros along a line, and then you met a rule that says, "Okay, I'm going to evolve this in time." And generally the simplest ones start with just each block of three ones and zeros have a rule that they will determinously go to either one or a zero, and you can actually classify all the different possibilities, a small number of possible cellular automata of that form. And what was discovered by various people, including Stephen Wolfram is some of these cellular automata have the feature that you start from almost nothing like 0, 0, 0, 0, 1, 0, 0, 0, 0, and you let it rip and it becomes wildly complex. Okay, so this is very provocative, very interesting. It's also not how physics works at all because as we said, physics conserves information. You can go forward or backwards. These cellular automata do not, they're not reversible in any sense. You've built in an arrow of time, you have a starting point, and then you evolve. So what I'm interested in is seeing how in the real world with the real laws of physics and underlying reversibility, but macroscopic irreversibility from entropy and the arrow of time, et cetera, how does that lead to complexity? I think that that's an answerable question. I don't think that cellular automata are really helping us in that one.

Lex Fridman
So what does the landscape of entropy in the universe look like?

Sean Carroll
Well, entropy is hard to localize. It's a property of systems, not of parts of systems. Having said that, we can do approximate answers to the question. The answer is black holes are huge in entropy. Let's put it this way, the whole observable universe that we were in had a certain amount of entropy before stars and planets and black holes started to form, 10 to the 88th. I can even tell you the number. Okay. The single black hole at the center of our galaxy has entropy, 10 to the 90. Single black hole at the of our galaxy has more entropy than the whole universe used to have not too long ago. So most of the entropy in the universe today is in the form of black holes.

Lex Fridman
Okay, that's fascinating first of all. But second of all, if we take black holes away, what are the different interesting perturbations in entropy across space? Where do we earthlings fit into that?

Sean Carroll
The interesting thing to me is that if you start with a system that is isolated from the rest of the universe and you start it at low entropy, there's almost a theorem that says if you're very, very, very low entropy, then the system looks pretty simple. Because low entropy means there's only a small number of ways that you can rearrange the parts to look like that. So if there's not that many ways, the answer's going to look simple. But there's also almost a theorem that says when you're at maximum entropy, the system is going to look simple because it's all smeared out. If it had interesting structure, then it would be complicated. So entropy in this isolated system only goes up. That's the second law of thermodynamics. But complexity starts low, goes up, and then goes down again. Sometimes people think that complexity or life or whatever is fighting against the second law of thermodynamics, fighting against the increase of entropy. That is precisely the wrong way to think about it. We are surfers riding the wave of increasing entropy. We rely on increasing entropy to survive. That is part of what makes us special. This table maintains its stability mechanically, which I mean there's molecules there, have forces on each other, and it holds up. You and I aren't like that. We maintain our stability dynamically by ingesting food, fuel, food, and water and air and so forth, burning it, increasing its entropy. We are non equilibrium, quasi steady-state systems. We are using the fuel the universe gives us in the form of low entropy energy to maintain our stability.

Lex Fridman
I just wonder what that mechanism of surfing looks like. First of all, one question to ask, do you think it's possible to have a kind of size of complexity where you have very precise ways or clearly defined ways of measuring complexity?

Sean Carroll
I think it is, and I think we don't. It's possible to have it, I don't think we yet have it because in part because complexity is not a univalent thing. There's different ideas that go under the rubric of complexity. One version is just [inaudible 01:56:41] complexity. If you have a configuration or a string of numbers or whatever, can you compress it so that you have a small program that will output that? That's [inaudible 01:56:51] complexity, but that's the complexity of a string of numbers. It's not like the complexity of a problem, computational complexity, the traveling salesman problem or factoring large numbers. That's a whole different kind of question that is also about complexity. So we don't have that sort of unified view of it.

Lex Fridman
So you think it's possible to have a complexity of a physical system?

Sean Carroll
Yeah, absolutely.

Lex Fridman
In the same way we do entropy?

Sean Carroll
Yeah.

Lex Fridman
You think that's a Sean Carroll paper or what?

Sean Carroll
We are working on various things. The glib thing that I'm trying to work on right now with a student is Complexo Genesis. How does complexity come to be if all the universe is doing is moving from low entropy to high entropy?

Lex Fridman
It's a sexy name.

Sean Carroll
It's a good name. Yeah, I like the name. I've just got to write the paper.

Lex Fridman
Sometimes a name, a rose by any other name. In which context, the birth of complexity are you most interested in?

Sean Carroll
Well, I think it comes in stages. So I think that if you go from ... I'm again a physicist, so biologists studying evolution will talk about how complexity evolves all the time, the complexity of the genome, the complexity of our physiology. But they take for granted that life already existed and entropy is increasing and so forth. I want to go back to the beginning and say the early universe was simple and low entropy and entropy increases with time, and the universe sort of differentiates and becomes more complex. But that statement, which is indisputably true, has different meanings because complexity has different meanings. So sort of the most basic primal version of complexity is what you might think of as configurational complexity. That's what [inaudible 01:58:39] gets at. How much information do you need to specify the configuration of the system? Then there's a whole other step where subsystems of the universe start burning fuel. So in many ways, a planet and a star

are not that different in configurational complexity. They're both spheres with density high at the middle and getting less as you go out. But there's something fundamentally different because the star only survives as long as it has fuel. I mean, then it turns into a brown dwarf or white dwarf for whatever. But as a star, as a main sequence star, it is an out of equilibrium system, but it's more or less static. If I spill the coffee mug and it falls, in the process of falling it' out of equilibrium, but it's also changing all the time. A specific kind of system is where it looks sort of macroscopically stationary, like a star, but underneath the hood, it's burning fuel to beat the band in order to maintain that stability. So as stars form, that's a different kind of complexity that comes to be. Then there's another kind of complexity that comes to be, roughly speaking at the origin of life, because that's where you have information really being gathered and utilized by subsystems of the universe. And then arguably, there's any number of stages past that. I mean, one of the most obvious ones to me is we talk about simulation theory, but you and I run simulations in our heads. They're just not that good. But we imagine different hypothetical futures. Bacteria don't do that. So that's the kind of information processing that is a form of complexity, and so I would like to understand all these stages and how they fit together.

Lex Fridman
Yeah, imagination.

Sean Carroll
Yeah, mental time travel.

Lex Fridman
Yeah. The things going on in my head when I'm imagining worlds are super compressed representations of those worlds, but [inaudible 02:00:32] get to the essence of them, and maybe it's possible with non-human computing type devices to do those kinds of simulations in more and more compressed ways.

Sean Carroll
There's an argument to be made that literally what separates human beings from other species on Earth is our ability to imagine counterfactual hypothetical futures.

Lex Fridman
Yeah, I mean, that's one of the big features. I don't know if it's a-

Sean Carroll
Everyone has their own favorite little feature, but that's why I said there's an argument to be made. I did a podcast episode on it with Adam Bulley. It developed slowly. I did a different podcast. Sorry to keep mentioning podcast episodes I did. But Malcolm Maciver, who is an engineer at Northwestern, has a theory about one of the major stages in evolution is when fish first climbed on the land. And I mean, of course that is a major stage of evolution, but in

particular, there's a cognitive shift because when you're a fish swimming under the water, the attenuation length of light in water is not that long. You can't see kilometers away. You can see meters away, and you're moving at meters per second. So all of the evolutionary optimization is make all of your decisions on a timescale of less than a second. When you see something new, you have to make a rapid fire decision what to do about it. As soon as you climb onto land, you can essentially see forever, you can see stars in the sky. So now a whole new mode of reasoning opens up where you see something far away and rather than saying, "Look up [inaudible 02:02:06]," I see this, I react. You can say, "Okay, I see that thing. What if I did this? What if I did that? What if I did something different?" And that's the birth of imagination eventually.

Lex Fridman
You've been critical on panpsychism.

Sean Carroll
Yes, you've noticed that.

Lex Fridman
Can you make the case for Panpsychism and against it? So panpsychism is the idea that consciousness permeates all matter. Maybe it's a fundamental force or a physics of the fabric of the universe.

Sean Carroll
Panpsychism, thought everywhere, consciousness everywhere.

Lex Fridman
To a point of entertainment, the idea frustrates you, which sort of as a fan is wonderful to watch, and you've had great episodes with panpsychists on your podcast where you go at it.

Sean Carroll
I had David Chalmers, who's one of the world's great philosophers, and he is panpsychism curious. He doesn't commit to anything, but he's certainly willing to entertain it. Philip Goff, who I've had, who is a great guy, but he's devoted to panpsychism. In fact, he's almost single-handedly responsible for the upsurge of interest in panpsychism in the popular imagination. And the argument for it is supposed to be that there is something fundamentally uncapturable about conscious awareness by physical behavior of atoms and molecules. So the panpsychist will say, "Look, you can tell me maybe someday, through advances of neuroscience and what have you, exactly what happens in your brain and how that translates into thought and speech and action. What you can't tell me is what it is like to be me. You can't tell me what I am experiencing when I see something that is red or that I taste something that is sweet. You can tell me what neurons fire, but you can't tell me what I'm experiencing, that first-person, inner subjective experience is simply not capturable by

physics." And therefore, this is an old argument, of course, but then therefore is supposed to be, I need something that is not contained within physics to account for that, and I'm just going to call it mind. We don't know what it is yet. We're going to call it mind, and it has to be separate from physics. And then there's two ways to go. If you buy that much, you can either say, okay, I'm going to be a dualist. I'm going to believe that there's matter and mind, and they're separate from each other and they're interacting somehow. Or that's a little bit complicated and sketchy as far as physics is going to go. So I'm going to believe in mind, but I'm going to put it prior to matter. I'm going to believe that mind comes first, and that consciousness is the fundamental aspect of reality and everything else, including matter and physics comes from it. That would be at least as simple as physics comes first. Now, the physicalist such as myself will say, I don't have any problem explaining what it's like to be you or what you experience when you see red. It's a certain way of talking about the atoms and the neurons, et cetera, that make up you. Just like the hardness or the brownness of this table, these are words that we attach to certain underlying configurations of ordinary physical matter. Likewise, sadness and redness or whatever are words that we attach to you to describe what you're doing. And when it comes to consciousness in general, I'm very quick to say I do not claim to have any special insight on how consciousness works other than I see no reason to change the laws of physics to account for it.

Lex Fridman
If you don't have to change the laws of physics, where do you think it emerges from? Is consciousness an illusion that's almost like a shorthand that we humans use to describe a certain kind of feeling we have when interacting with the world, or is there some big leap that happens at some stage?

Sean Carroll
I almost never use the word illusion. Illusion means that there's something that you think you're perceiving that is actually not there. Like an oasis in the desert is an illusion. It has no causal efficacy. If you walk up to where the oasis is supposed to be, you'll say you were wrong about it being there. That's different than something being emergent or non-fundamental, but also real. This table is real, even though I know it's made of atoms, that doesn't remove the realness from the table. I think that consciousness and free will and things like that are just as real in tables and chairs.

Lex Fridman
Oasis in the desert does have causal efficacy in that you're thirsty [inaudible 02:06:53].

Sean Carroll
It leads you to draw incorrect conclusions about the world.

Lex Fridman
Sure, but imagining a thing can sometimes bring it to reality, as we've seen, and that has a kind of causal efficacy.

Sean Carroll
But your understanding of the world in a way that gives you power over it and influence over it is decreased rather than increased by believing in that oasis. That is not true about consciousness or this table.

Lex Fridman
You don't think you can increase the chance of a thing existing by imagining it existing?

Sean Carroll
No. Unless you build it or make it.

Lex Fridman
No, that's what I mean. Imagining humans can fly if you're the Wright brothers.

Sean Carroll
[inaudible 02:07:37] imagine that humans are flying, in terms of counterfactuals in the future, absolutely. Imagination is crucially important, but that's not an illusion. That's just a imagination.

Lex Fridman
Okay. The possibility of the future versus what the reality is. I mean, the future is a concept, so you can ... Time is just a concept, so you can play with that. Time is just a concept so you can play with that. But yes, reality. So, to you ... So for example, I love asking this. So, Donald Hoffman thinks that the entirety of the conversation we've been having about space-time is an illusion. Is it possible for you to steelman the case for that? Can you make the case for and against reality, as I think he writes, that the laws of physics as we know them with space-time, is it interface to a much deeper thing that we don't at all understand and that we're fooling ourselves by constructing this world?

Sean Carroll
Well, I think there's part of that idea that is perfectly respectable and part of it that is perfectly nonsensical and I'm not even going to try to steelman the nonsensical part. The real part to me is what is called structural realism, so we don't know what the world is at a deep fundamental level. Let's put ourselves in the minds of people living 200 years ago, they didn't know about quantum mechanics, they didn't know about relativity, that doesn't mean they were wrong about the universe that they understood, they had Newton's laws, they could predict what time the sun was going to rise perfectly well. In the progress of science, the words that would be used to give the most fundamental description of how you were

predicting the sun would rise changed because now you have curved space-time and things like that and you didn't have any of those words 200 years ago. But the prediction is the same, why? Because that prediction, independent of what we thought the fundamental ontology was, the prediction pointed to something true about our understanding of reality. To call it an illusion is just wrong, I think. We might not know what the best, most comprehensive way of stating it is but it's still true.

Lex Fridman
Is it true in the way, for example, belief in God is true? Because for most of human history, people have believed in a God or multiple gods and that seemed very true to them as an explanation for the way the world is, some of the deeper questions about life itself with the human condition and why certain things happen, that was a good explainer. So, to you, that's not an illusion?

Sean Carroll
No, I think that was completely an illusion. I think it was a very, very reasonable illusion to be under. There are illusions, there are substantive claims about the world that go beyond predictions that we can make and verify which later turned out to be wrong and the existence of God was one of them. If those people at that time had abandoned their belief in God and replaced it with a mechanistic universe, they would've done just as well at understanding things. Again, because there are so many things they didn't understand, it was very reasonable for them to have that belief, it wasn't that they were dummies or anything like that. But that is, as we understand the universe better and better, some things stick with us, some things get replaced.

Lex Fridman
So, like you said, you are a believer of the mechanistic universe, you're a naturalist and, as you've described, a poetic naturalist.

Sean Carroll
That's right.

Lex Fridman
What's the word poetic ... What is naturalism and what is poetic naturalism?

Sean Carroll
Naturalism is just the idea that all that exists is the natural world, there's no supernatural world. You can have arguments about what that means but I would claim that the argument should be about what the word supernatural means, not the word natural. The natural world is the world that we learn about by doing science. The poetic part means that you shouldn't be too, I want to say, fundamentalist about what the natural world is. As we went from Newtonian space-time to Einsteinian space-time, something is maintained there, there is a

different story that we can tell about the world. And that story, in the Newtonian regime, if you want to fly a rocket to the moon, you don't use general relativity, you use Newtonian mechanics, that story works perfectly well. The poetic aspect of the story is that there are many ways of talking about the natural world and, as long as those ways latch onto something real and causally efficacious about the functioning of the world, then we attribute some reality and truth to them.

Lex Fridman
So, the poetic really looks at the, let's say, the pothead questions at the edge of science is more open to them.

Sean Carroll
It's doing double duty a little bit so that's why it's confusing. The more obvious respectable duty it's doing is that tables are real. Even though you know that it's really a quantum field theory wave function, tables are still real, there are a different way of talking about the underlying deeper reality of it. The other duty it's doing is that we move beyond purely descriptive vocabularies for discussing the universe onto normative and prescriptive and judgmental ways of talking about the universe. This painting is beautiful, that one is ugly. This action is morally right, that one is morally wrong. These are also ways of talking about the universe, they are not fixed by the phenomena, they're not determined by our observations, they cannot be ruled out by a crucial experiment but they're still valid. They might not be universal, they might be subjective but they're not arbitrary and they do have a role in describing how the world works.

Lex Fridman
So, you don't think it's possible to construct experiments that explore the realms of morality and even meaning? So, those are subjective?

Sean Carroll
Yeah. They're human, they're personal.

Lex Fridman
But do you think that's just because we don't have a ... The tools of science have not expanded enough to incorporate the human experience?

Sean Carroll
No, I don't think that's what it is. I think that what we mean by aesthetics or morality are we're attaching categories, properties to things that happen in the physical world and there is always going to be some subjectivity to our attachment and how we do that and that's okay and, the faster we recognize that and deal with it, the better off we'll be.

Lex Fridman
But if we deeply and fully understand the function of the human mind, it won't be able to incorporate that?

Sean Carroll
No. That will absolutely be helpful in explaining why certain people have certain moral beliefs, it won't justify those beliefs as right or wrong.

Lex Fridman
Do you think it's possible to have a general relativity that includes the observer effect where the human mind is the observer?

Sean Carroll
Sure.

Lex Fridman
How we morph in the same way gravity morphs space-time, how does the human mind morph reality and have a very thorough theory of how that morphing actually happens?

Sean Carroll
That's a very pothead question, Lex, but-

Lex Fridman
I'm sorry.

Sean Carroll
It's okay.

Lex Fridman
But do you think it's possible?

Sean Carroll
The answer is yes. I think that there's no-

Lex Fridman
Okay, all right.

Sean Carroll
I think we are part of the physical world, the natural world. Physicalism would've been just as good a word to use as naturalism, maybe even a more accurate word but it's a little bit more off-putting so I do want to snap your more attractive label than physicalism.

Lex Fridman
Are there limits to science?

Sean Carroll
Sure. We just talked about one, right? Science can't tell you right from wrong. You need science to implement your ideas about right and wrong. If you are functioning on the basis of an incorrect view of how the world works, you might very well think you're doing right but actually be doing wrong but all the science in the world won't tell you which action is right and which action is wrong.

Lex Fridman
Dictators and people in power sometimes use science as an authority to convince you what's right and wrong, studying Nazi science is fascinating.

Sean Carroll
Yeah. But there's an instrumentalist view here, you have to first decide what your goals are and then science can help you achieve those goals. If your goals are horrible, science has no problem helping you achieve them, science is happy to help out.

Lex Fridman
Let me ask you about the method behind the madness on several aspects of your life. So, you mentioned your approach to writing for research and writing popular books, how do you find the time of the day? What's the day in the life of Sean Carroll looks like?

Sean Carroll
Very unclear how I have the time, honestly.

Lex Fridman
So, you don't have a thing where, in the morning, you try to fight for two hours somewhere?

Sean Carroll
I don't, I'm really terrible at that. My strategy for finding time is just to ignore interruptions and emails but it's a different time every day, some days it never happens, some weeks it never happens.

Lex Fridman
Oh, really? You're able to pull it off? Because you're extremely prolific. So, you're able to have days where you don't write-

Sean Carroll
Oh, my god, yes. Yeah.

Lex Fridman
... and still write the next day?

Sean Carroll
Right.

Lex Fridman
Oh, wow. That's a rare thing, right? A lot of prolific writers will-

Sean Carroll
I guess it's true.

Lex Fridman
... carve out two hours because, otherwise, it just disappears.

Sean Carroll
Right. No, I get that. Yeah, I do. And yeah, it just everyone has their foibles or whatever so I'm not able to do that, therefore, I have to just figure it out on the fly.

Lex Fridman
And what's the actual process look like when you're writing popular stuff? You get behind a computer?

Sean Carroll
Yeah, get behind a computer. My way of doing it ... So, my wife, Jennifer, is a science writer but it's interesting because our techniques are entirely different. She will think about something but then she'll free write, she'll just sit at a computer and write I think this, I think this, I think this. And then that will be vastly compressed, edited, rewritten or whatever until the final thing happens. I will just sit there silently thinking for a very long time and then I'll write what is almost the final draft. So, a lot of it happens. There might be some scribbles for an outline or something like that but a lot of it is in my brain before it's on the page.

Lex Fridman
So, that's the case for The Biggest Ideas in the Universe, the quanta book and the space, time and motion book?

Sean Carroll
Yeah, Quanta and Fields, which is actually mostly about quantum field theory and particle physics, that's coming out in May. And that is I'm letting people in on things that no other book lets them in on so I hope it's worth it. It's a challenge because it's a lot of equations.

Lex Fridman
You did the same thing with Space, Time and Motion. You did something quite interesting which is you made the equation the centerpiece of a book.

Sean Carroll
Right, there's a lot of equations. Book two goes further in those directions than book one did. So, it's more cool stuff, it's also more mind-bending, it's more of a challenge. Book three that I'm writing right now is called Complexity and Emergence.

Lex Fridman
Oh wow.

Sean Carroll
And that'll be the final part of the trilogy.

Lex Fridman
Oh, that's fascinating. So, there's a lot of, probably, ideas there, that's a real cutting edge.

Sean Carroll
Well, but I'm not trying to be cutting edge. In other words, I'm not trying to speculate in these books. Obviously, in other books, I've been very free about speculating but the point of these books is to say things that, 500 years from now, will still be true. And so, there are some things we know about complexity and emergence and I want to focus on those. And I will mention, I'm happy to say, this is something that needs to be speculated about but I won't pretend to be telling you what one is the right one.

Lex Fridman
You somehow found the balance between the rigor of mathematics and still accessible which is interesting.

Sean Carroll
I try. Look, these three books, the Biggest Ideas books are absolutely an experiment. They're going to appeal to a smaller audience than other books will but that audience should love them. My 19-year-old Self would've been so happy to get these books, I can't tell you.

Lex Fridman
Yeah, in terms of looking back in history, those are books ... The trilogy would be truly special in that way.

Sean Carroll
Worked for Lord of the Rings so I figured why not me.

Lex Fridman
You and Tolkien.

Sean Carroll
Yeah.

Lex Fridman
Just different styles, different topics.

Sean Carroll
Same ultimate reality.

Lex Fridman
We mentioned Mindscape Podcast, I love it. You interview a huge variety of experts from all kinds of fields so just several questions I want to ask. How do you prepare? How prepare to have a good conversation? How do you prepare in a way that satisfies, makes your own curious mind happy, all that kind of stuff?

Sean Carroll
Yeah, no, these are great questions and I've struggled and changed my techniques over the years, it's a over five-year-old podcast, might be approaching six years old now. I started out over-preparing when I first started, I had a journey that I was going to go down. Many of the people I talk to are academics or thinkers who write books so they have a story to tell, I could just say, "Okay, give me your lecture and then, an hour later, stop." So, the mistake is to anticipate what the lecture would be and to ask the leading questions that would pull it out of them. What I do now is much more here are the points, here are the big questions that I'm interested in and so I have a much sketchier outline to start and then try to make it more of a real conversation. I'm helped by the fact that it is not my day job so I strictly limit myself to one day of my life per podcast episode on average, some days take more. And that includes, not just doing the research, but inviting the guest, recording it, editing it, publishing it. So, I need to be very, very efficient at that, yeah.

Lex Fridman
You enforce constraints for yourself in which creativity can emerge.

Sean Carroll
That's right, that's right. And look, sometimes, if I'm interviewing a theoretical physicist, I can just go in. And where I'm interviewing an economist or a historian, I have to do a lot of work.

Lex Fridman
Do you ever find yourself getting lost in rabbit holes that serve no purpose except satisfying your own curiosity and then potentially expanding the range of things you know that can help your actual work and research and writing?

Sean Carroll
Yes, on both counts. Some people have so many things to talk about that you don't know where to start or finish, others have a message. And one of the thing I discovered over the course of these years is the correlation with age. There are brilliant people and I try very hard on the podcast to get all sorts of people, different ages and things like that and, bless their hearts, the most brilliant young people are not as practiced at wandering past their literal research. The have less mastery over the field as a whole, much less how to talk about it. Whereas, certain older people just have their pad answers and that's boring. So, you want somewhere in between, the ideal person who has a broad enough of a scope that they can wander outside their specific papers they've written but they're not overly practiced so they're just giving you their canned answers.

Lex Fridman
I feel like there's a connection to the metaphor of entropy and complexity, as you said there.

Sean Carroll
Yeah. Edge of chaos, yeah.

Lex Fridman
You also do incredible AMAs and people should sign up to your Patreon because you can get to ask questions, Sean Carroll. Well, for several hours, you just answer in fascinating ways some really interesting questions. Is there something you could say about the process of finding the answers to those?

Sean Carroll
That's a great one. Again, it's evolved over time. So, the Ask me Anything episodes were first, when I started doing them, they were only for Patreon subscribers to both listen to and to ask the questions. But then I actually asked my Patreon subscribers, "Would you like me to release them publicly?" and they overwhelmingly voted yes so I do that. So, the Patreon supporters ask the questions, everyone can listen. And also, at some point, I really used to try to answer every question but now there's just too many so I have to pick and that's fraught with peril and my personal standard for picking questions to answer is what are the ones I think I have interesting answers to give for. So, that both means, if it's the same old question about special relativity that I've gotten a hundred times before, I'm not going to answer it because you can just Google that, it's easier. There are some very clear attempts to ask an interesting question that, honestly, I don't have an answer to. Like, " I read this science fiction novel, what do you think about it?" I'm like, "Well, I haven't read it so I can't

help you there." "What's your favorite color?" "I could tell you what it is but it's not that interesting." And so, I try to make it a mix, I try to ... It's not all physics questions, not all philosophy questions, I will talk about food or movies or politics or religion if that's what people want to. I keep suggesting that people ask me for relationship advice but they never do.

Lex Fridman
Yeah, I don't think I've heard one.

Sean Carroll
Yeah, I'm willing to do it. I'm a little reluctant because I don't actually like giving advice but I'm happy to talk about those topics. I want to give several hours of talking and I want to try to say things that I haven't said before and keep it interesting, keep it rolling. If you like this question, wait for the next one.

Lex Fridman
What are some of the harder questions you've gotten? Do you remember? What kinds of questions are difficult for you?

Sean Carroll
Rarely but occasionally people will ask me a super insightful philosophy question. I hadn't thought of it things in exactly that way and I try to recognize that. A lot of times, it is the opposite where it's like, "Okay, you're clearly confused and I'm going to try to explain the question you should have asked."

Lex Fridman
I love those. Yeah, why that's the wrong question or that kind of stuff, that's great.

Sean Carroll
Right.

Lex Fridman
That's great.

Sean Carroll
But the hard questions, I don't know. I don't actually answer personal questions very much. The most personal I will get are questions like what do you think of Baltimore, that much I can talk about. Or how are your cats doing, happy to talk about the cats in infinite detail. But very personal questions I don't get into.

Lex Fridman
But you even touch politics and stuff like this.

Sean Carroll
Yeah, no, very happy to talk about politics. I try to be clear on what is professional expertise, what is just me babbling, what is my level of credence in different things, where you're allowed to disagree, whether, if you disagree, you're just wrong and people can disagree with that also. But I do think I'm happy to go out on a limb a little bit, I'm happy to say, "Look, I don't know but here's my guess." I just did a whole solo podcast which was exactly that. And it's interesting, some people are like, "Oh, this was great," and there's a whole bunch of people who are like, "Why are you talking about this thing that you are not the world's expert in?"

Lex Fridman
Well, I love the actual dance between humility and having a strong opinion on stuff, it's a fascinating dance to pull off. And I guess the way to do that is to just expand into all kinds of topics and play with ideas and then change your mind and all that kind of stuff.

Sean Carroll
Yeah, it is interesting because, when people react against you by saying you are being arrogant about this, 99.999% of the time, all they mean is I disagree. That's all they really mean, right?

Lex Fridman
Yeah.

Sean Carroll
At a very basic level, people will accuse atheists of being arrogant and I'm like, "You think God exists and loves you and you're telling me that I'm arrogant?" I think that all of this is to say just advice. When you disagree with somebody, try to specify the substantive disagreement, try not to psychologize them. Try to say, "Oh, you're saying this because of this." Maybe it's true, maybe you're right. But if you had an actual response to what they were saying, that would be much more interesting.

Lex Fridman
Yeah, I wonder why it's difficult for people to say or to imply I respect you, I like you but I disagree on this and here's why I disagree. I wonder why they go to this place of, well, you're an idiot or you're egotistical or you're confused or you're naive or you're all the kinds of words as opposed to I respect you as a fellow human being exploring the world of mysteries all around us and I disagree.

Sean Carroll
I will complicate the question even more because there's some people I don't respect or like. And I once read a blog post, I think it was called The Grid of Disputation and I had a two by two grid and it's are you someone I agree with or disagree with, are you someone who I

respect or don't and all four quadrants are very populated. So, what that means is there are people who I like and I disagree with and there are people who agree with me and I have no respect for at all, the embarrassing allies quadrant, that was everyone's favorite.

Lex Fridman
That's great.

Sean Carroll
So, I just think being honest, trying to be honest about where people are. But if you actually want to move a conversation forward, forget about whether you like or don't like somebody, explain the disagreement, explain the agreement. But you're absolutely right, I completely agree, as a society, we are not very good at disagreeing, we instantly go to the insults.

Lex Fridman
Yeah. And even on a deeper level, I think, at some deep level, I respect and love the humanity in the other person.

Sean Carroll
Yup.

Lex Fridman
You said that general relativity is the most beautiful theory ever.

Sean Carroll
So far.

Lex Fridman
What do you find beautiful about it?

Sean Carroll
Let's put it this way. When I teach courses, there's no more satisfying subject to teach than general relativity and the reason why is because it starts from very clear, precisely articulated assumptions and it goes so far. And when I give my talk, you can find it online, I'm probably not going to give it again, the book one of the Biggest Ideas talk was building up from you don't know any math or physics, an hour later, you know Einstein's equation for general relativity. And the punchline is the equation is much smarter than Albert Einstein because Albert Einstein did not know about the Big Bang, he didn't know about gravitational waves, he didn't know about black holes but his equation did. And that's a miraculous aspect of science more generally but general relativity is where it manifests itself in the most absolutely obvious way.

Lex Fridman
A human question, what do you think of the fact that Einstein didn't get the Nobel Prize for general relativity?

Sean Carroll
Tragedy. He should have gotten maybe four Nobel Prizes, honestly. He certainly should have got-

Lex Fridman
That and what?

Sean Carroll
The photoelectric effect was 100% worth the Nobel Prize because, and people don't quite get this, who cares about the photoelectric effect, that's this very minor effect. The point is his explanation for the photoelectric effect invented something called the photon, that's worth the Nobel Prize. Max Planck gets credit for this in 1900 explaining black-body radiation by saying that, when a little electron is jiggling in a object at some temperature, it gives off radiation in discrete chunks rather than continuously. He didn't quite say that's because radiation is discrete chunks. It's like having a coffee maker that makes one cup of coffee at a time, it doesn't mean that liquid comes in one cup quanta, it's just that you are dispensing it like that. It was Einstein in 1905 who said light is quanta and that was a radical thing. So, clearly, that was not a mistake. But also special relativity clearly deserved the Nobel Prize and general relativity clearly deserved the Nobel Prize. Not only were they brilliant but they were experimentally verified, everything you want.

Lex Fridman
So, separately you think?

Sean Carroll
Yeah. Yeah, absolutely.

Lex Fridman
Oh, humans.

Sean Carroll
Yeah.

Lex Fridman
Whatever the explanation there.

Sean Carroll
Edwin Hubble never won the Nobel Prize for finding the universe was expanding.

Lex Fridman
Yeah. And even the fact that we give prizes is almost silly and we limit the number of people that get the prize and all that.

Sean Carroll
I think that Nobel Prize has enormous problems. I think it's probably a net good for the world because it brings attention to good science. I think it's probably a net negative for science because it makes people want to win the Nobel Prize.

Lex Fridman
Yeah, there's a lot of fascinating human stories underneath it all. Science is its own thing but it's also a collection of humans and it's a beautiful collection. There's tension, there's competition, there's jealousy but there's also great collaborations and all that kind of stuff. Daniel Kahneman, who recently passed, is one of the great stories of collaboration in science.

Sean Carroll
Yeah, [inaudible 02:34:01].

Lex Fridman
So, all of it, all of it, that's what humans do. And Sean, thank you for being the person that makes us celebrate science and fall in love with all of these beautiful ideas in science, for writing amazing books, for being legit and still pushing forward the research science side of it and for allowing me and these pothead questions and also for educating everybody through your own podcast. Everybody should stop everything and subscribe and listen to every single episode of Mindscape. So, thank you, I've been a huge fan forever, I'm really honored that you would speak with me in the early days when I was still starting this podcast in Meanings of the World.

Sean Carroll
I appreciate it. Thanks very much for having me on. Now that you're a big deal, still having me on.

Lex Fridman
Thank you, Sean. Thanks for listening to this conversation with Sean Carroll. To support this podcast, please check out our sponsors in the description. And now, let me leave you with some words from Richard Feynman. Study hard what interests you the most in the most undisciplined, irreverent and original manner possible. Thank you for listening and hope to see you next time.