SAM HARRIS

THE BLOG

The Multiverse & You (& You & You & You...)

Podcast Transcript

Book News | Neuroscience | Philosophy | Physics | Science | April 20, 2016

(Photo via DaveMorrowPhotography.com)

Known as “Mad Max” for his unorthodox ideas and passion for adventure, Max Tegmark’s scientific interests range from precision cosmology to the ultimate nature of reality, all explored in his new popular book *Our Mathematical Universe.* Tegmark is a professor of physics who has published more than two hundred technical papers and been featured in dozens of science documentaries. His work with the Sloan Digital Sky Survey on galaxy clustering shared the first prize in *Science* magazine’s “Breakthrough of the Year: 2003.” For more information about his work, please visit his MIT website and the Future of Life Institute.

* * *

Welcome to the Waking Up podcast. This is Sam Harris. Today I’ll be speaking with Max Tegmark. Max is a physicist at MIT, and a cosmologist, in particular. He’s published over 200 technical papers, and he’s been featured in dozens of science documentaries. He’s now an increasingly influential voice on the topic of artificial intelligence, because his Future of Life Institute deals with this and other potential existential threats.

Max has written one book for the general reader which I found incredibly valuable, entitled *Our Mathematical Universe.* We’ll be talking about some of that today. I really enjoyed speaking with Max. We discussed the foundations of science and what distinguishes science from non-science. We talked about the mysterious utility of mathematics in the natural sciences.

We also talked, for quite some time, about our current picture of the universe from a cosmological perspective, which opens onto the fascinating and totally counterintuitive concept of the multiverse—which, as you’ll see, entails the claim that there may well be a functionally infinite number of people just like yourself leading nearly identical lives and every other possible life at this moment elsewhere in the universe. This is my candidate for the strangest idea that is still scientifically plausible.

And finally, we talked about the dangers of advances in artificial intelligence. In any case, Max is a fascinating guy, and I hope you enjoy our conversation. I also hope you’ll buy his book, because it is well worth reading. And now I give you Max Tegmark…

SH: How are you doing, Max? Thanks for coming on the podcast.

MT: Thank you for having me. It’s great to be on.

SH: It’s really a pleasure to talk to you. I have a lot I want to talk about, including your book, *Our Mathematical Universe,* which I highly recommend to our listeners. I’m going to talk about some of what I find most interesting in that book, but it by no means exhausts the contents.

I really consider your book a huge achievement. You’ve managed to make an up-to-the-minute picture of the state of physics—and cosmology, in particular—truly accessible to a general reader. That’s certainly not something that all your colleagues can claim to have achieved. So congratulations on that.

MT: Thank you for your kind words. It’s important to remember also, of course, that if in speaking about these things, or reading my book, one feels that one doesn’t understand quite everything about our cosmos, you know, nobody else does either. So that’s quite okay. And in fact, that’s really very much part of the charm of studying the cosmos—that we still have these great mysteries that we can ponder.

SH: Indeed. And I’m going to drive rather directly toward those mysteries. But first I want to give some context here. You and I met in San Juan, Puerto Rico, at a conference you helped organize on the frontiers of artificial intelligence research—in particular, the emerging safety concerns there. I hope we'll eventually get to that, because our shared interest is in AI at the moment. But I do want to talk first about physics first, and then we can discuss the armies of lethal robots that may await us.

MT: Sounds great. It seems pretty clear to me, from our conversations, that we also have a very strong shared interest in looking at this reality out there and pondering what its true nature really is.

SH: Let's start there—at the foundations of our knowledge and the foundations of science. As you know, in science, we are making our best effort to arrive at a unified understanding of reality. I think there are many people in our culture, many in humanities departments, who think that no such understanding is possible. They think no view of the world can encompass subatomic particles and cocktail parties and everything in between.

But I think that from the point of view of science, we have to believe that there is. We may use different concepts at different scales, but there shouldn't be radical discontinuities between different scales in our understanding of reality. I trust that this is an intuition you share.

MT: Yes. When someone says that they think reality is just a social construct or whatnot, then other people get upset and say, if you think gravity is a social construct, I encourage you to take a step out through my window here on the sixth floor. If you drill down into what this conflict comes from, it's just that they're using that R-word, "reality," in very different ways.

As a physicist, the way I use the word "reality" is, I assume that there is something out there independent of me as a human. I assume that the Andromeda galaxy would continue existing even if I weren't here, for example. And then we take this very humble approach to say, "Okay, if there is some stuff that exists out there—our physical reality, let's call it—let's look at it as closely as we can and try to figure out what properties it has." If there's some confusion about something, that's our problem and not reality's problem.

There's no doubt in my mind that our universe knows perfectly well what it's doing, and it functions in some way. We physicists have so far failed to figure out what that way is. We're in this schizophrenic situation where we can't even make quantum mechanics talk to relativity theory properly. But that's the way I see it. Simply a failure, so far, in our own creativity. Not only do I guess that there is a reality out there independent of us, but I actually feel it's quite arrogant to say the opposite.

SH: Right.

MT: Because it sort of presumes that we humans should go center stage. Solipsists say that there is no reality without themselves. Ostriches in the apocryphal story make the assumption that things that they don't see don't exist. But even very respected scientists go down this slippery slope sometimes.

Niels Bohr, one of the founders of quantum mechanics, famously said, “No reality without observation,” which sort of puts humans center stage and denies that there can be things you should call “reality” without us. But I think that’s very arrogant. I think we could use a good dose of humility. So my starting point is, there is something out there, and let’s try to figure out how it works.

SH: I think we’ll get to Bohr and to his Copenhagen interpretation of quantum mechanics at some point, at least on the fly, because as you probably know, it really is the darling interpretation of New Age philosophers and spiritualists. It’s something that I think we have reason to be skeptical about. But inconveniently for us, this skepticism about the possibility of understanding reality sneaks in the back door, somewhat paradoxically, by virtue of taking science—in particular, evolutionary biology—seriously.

And this is something you and I were talking about when we last met. I think at one point in the conversation, I observed (as almost everyone has who thinks about evolution) that one thing we can be sure of is that our cognitive capacities, our common sense, and our intuitions about reality have not evolved to equip us to understand reality at the smallest possible scale, or at the largest, or when things are moving incredibly fast, or when they are very old.

We have intuitions that are tuned for things at human scale, that are moving relatively slowly, so that we can decide whether we can mate with them or eat them or whether they’re going to eat us. You and I were talking about this, and I said that it’s no surprise, therefore, that the deliverances of science, in particular your area of science, are deeply counterintuitive.

MT: That’s right.

SH: You did me one better, however. You said that not only is it *not* surprising, it would be surprising—and, in fact, give you reason to mistrust your theories—if they *were* aligned with common sense. We should expect the answers at the back of the book of Nature to be deeply counterintuitive. I want you to expand on that a little bit.

MT: Yes, that’s exactly right. I think that’s a very clear prediction of Darwin’s ideas, if you take them seriously—that whatever the ultimate nature of reality is, it should seem really weird and counterintuitive to us. Because developing a brain advanced enough to understand new concepts is costly in evolution. We wouldn’t have evolved and spent a lot of energy increasing our metabolism, etc., if it didn’t help in any way.

If some cavewoman spent too much time pondering what was out there beyond all the stars that she could see, or subatomic particles, she might not have noticed the tiger that snuck up behind her and gone clean out of the gene pool. Moreover, this is not just a natural logical prediction, but it’s a testable prediction. Darwin lived a long time ago, right? And we can look at what has happened since then when we’ve used technology to probe things beyond what we could experience with our senses.

So the prediction is that whenever we, with technology, study physics that was inaccessible to our ancestors, it will seem weird. So let’s look at the fact sheet, at the scorecard. We’ve studied what happens when things go much faster than our ancestors could imagine, near the speed of light: Time slows down. *Whoa!* So weird that Einstein never even got the Nobel Prize for it, because my Swedish curmudgeonly countryman on the Nobel committee thought it was too weird.

You look at what happens when things are really, really large and you get black holes, which were considered so weird. Again, it took a long time until people started to accept them. And then you look at what happens when you make things really small—so small that our ancestors couldn't see them. And you find that elementary particles can be in several places at once. Extremely counterintuitive, to the point that people are still arguing about what it means, even though they all concede that particles really can do this weird stuff. And the list goes on.

Whenever you take any parameter out of the range of what our ancestors experienced, really weird things happen. If you have very high energies, for example—like when you smash two particles together near the speed of light with the Large Hadron Collider at CERN. Now, if you collide a proton and an anti-proton together, and out pops a Higgs boson, that's about as intuitive as if you collide a Volkswagen with an Audi and out comes a cruise ship. And yet, this is the way the world works.

So I think the verdict is in. Whatever the nature of reality actually is, it's going to seem really weird to us. And if we therefore dismiss physical theories just because they seem counterintuitive, we're almost certainly going to dismiss whatever the correct theory is once someone actually tells us about it.

SH: I'm wondering, though, whether this slippery slope is in fact more slippery than we're admitting here. How do we resist the slide into total epistemological skepticism? For instance, why trust our mathematical intuitions, or the mathematical concepts borne of them, or the picture of reality in physics that's arrived at through this bootstrapping of our intuitions into areas that are counterintuitive?

I understand why we should trust these things pragmatically—provided they seem to work. We can build airplanes that actually fly—and there's a difference between an airplane that flies and one that doesn't. But as a matter of epistemology, why should we trust the picture of reality that math allows us to bring into view if, again, we are just apes who have used the cognitive capacities that have evolved without any constraints that would have kept them in line with reality at large? After all, mathematics—insofar as we apprehend it, discover it, invent it—is just an extension of those very humble capacities.

MT: It's a very good question. And some people tell me sometimes that the theories that physicists discuss at conferences, from black holes to parallel universes, sound even crazier than a lot of myths from old time about flame-throwing dragons and whatnot, So shouldn't we dismiss the physics just as we dismiss these myths?

To me, there's a huge difference here, in that these physics theories, even though they sound crazy, actually make predictions that we can test. That is really the crux of it. If you take the theory of quantum mechanics seriously, for example, and assume that particles can be in several places at once, then you predict that you should be able to build this thing called a transistor, which you can combine in vast numbers and build this thing called a cell phone, and it actually works.

Good luck getting some useful technology using the fire dragon hypothesis or whatever. This is very linked, I think, to where we should draw the borderline between what's science and what's not science. Some people think that the line should go between that which seems intuitive and not crazy and that which feels too crazy.

I'm arguing against that—because black holes seemed very crazy at the time, and now we've found loads of them in the sky. To me, the line in the sand that divides science from what's not science is—the way I think about it is, what makes me a scientist is—that I would much rather have questions I can't answer than answers I can't question.

SH: One thing you're emphasizing here is that the distinction between science and non-science is not in the strangeness or the seeming acceptability of one's conclusions, but in the methodology by which one arrived at those conclusions. Of course, falsifiable predictions are part of that. I'm not sure you would say that a narrow Popperian conception of science—that is, the primacy of falsifiable claims—subsumes all of science. There are clearly scientifically coherent things we could say about the nature of reality that aren't falsifiable as a matter of practice. We know there's an answer there, and we just know that no one *has* the answer.

The very prosaic example I often use is, how many birds are in flight over the surface of the earth at this moment? We don't know. In fact, we know we'll never know, because the answer changed before I could get to the end of this sentence. But it was a totally coherent question to ask, and we know that it had an integer answer.

Leaving spooky quantum mechanics or parallel universes aside, if we're just talking about the earth and birds as objects, we can't get access to the data. But we know, in some basic sense, that this reality that extends beyond our perception guarantees that in principle, the data exist.

I think you say, at some point in your book, that a theory doesn't have to be testable across the board; just parts of it have to be testable to give some level of credence to its overall picture. Is that how you view it?

MT: I'm actually pretty sympathetic to Popper. The idea of testability works fine for even these crazy-sounding concepts like parallel universes and black holes, as long as we remember that what we test are theories—specific mathematical theories that we can write down. Right? Parallel universes are not a theory; they're a prediction from certain theories.

A black hole isn't a theory either; it's a prediction from Einstein's general relativity theory. And once you have a theory in physics, it's testable as long as it predicts at least one thing that you can check. Because then you can falsify it if you check that thing and it's wrong. Whereas, just because it happens to also make some other predictions for things you can never test, that doesn't make it non-scientific, as long as there is still something you can test.

SH: Yes. That's a much better construal of Popper. We falsify theories, not all conceivable predictions derived from them.

MT: Black holes, for example. The theory of general relativity predicts exactly what will happen to you if you fall into the monster black hole in the middle of a galaxy that weighs four million times as much as the sun. It predicts exactly when you'll get spaghettified and how, and so on. Except you can never actually do that experiment and then write an article about it, because you're inside the event horizon and the information can't come out.

Nonetheless, that's a testable theory. Because general relativity also predicts loads of other things, such as how your GPS works, which we can test with great precision. And when the theory passes a lot of tests for things that we can make, and we start to take the theory seriously, then I feel we have to be honest and also take seriously the other predictions from the theory, whether we like them or not.

We can't just cherry-pick and say, "Hey, I love what the general relativity theory does for GPS and the bending of light and the perihelion—the weird orbit of mercury and so on—but I don't really like the black holes, so I'm going to opt out of that prediction."

That you cannot do, the way you can say, "I want coffee" and opt out of the caffeine and buy decaf. In physics, once you buy the theory, you have to buy the whole product. And if you don't like any of the predictions, well, then you have to try to come up with a different mathematical theory, which doesn't have that prediction but still explains everything else. And that's often very hard.

People have tried for 100 years to do that with Einstein's gravity theory, to get rid of the black holes. And so far, they've pretty much failed. That's why people have been dragged kicking and screaming into believing in, or at least taking very seriously, black holes.

It's the same thing with these various kinds of parallel universes. It's precisely because people have tried so hard to come up with alternative theories that explain how to make computers—and that don't have these weird predictions and failed—that you're starting to take it more seriously.

SH: We're going to get to parallel universes, because that's really where, I think, people's intuitions break down entirely. But before we go there, I want to linger on this question about the primacy and strange utility of mathematics. At one point in your book, you cite the oft-cited paper by Wigner—I think he wrote it in the '60s—entitled "The Unreasonable Effectiveness of Mathematics in the Natural Sciences." This is something that many scientists have remarked on.

There seems to be a mysterious property of these abstract structures and chains of reasoning where mathematics is uniquely useful for describing the physical world and making predictions about things that you would never anticipate, but for the fact that the mathematics is suggesting that these things should exist.

MT: That's right.

SH: This has lured many scientists into, essentially, mysticism or, at the very least, philosophical Platonism—sometimes even religion. Positing mathematical structures or even pure mathematical concepts like numbers, that exist in almost a Platonic state beyond the human mind. I'm wondering if you share some of that mathematical idealism.

I also wanted to get your reaction to an idea that I believe I got from a cognitive scientist who died, I think in the '40s, Kenneth Craik, who published a book in 1943 where he anticipates Wigner by about 20 years. In passing, he tried to resolve this mystery about the strange utility of mathematics. He speculated that there must be some isomorphism between brain processes that represent the physical world and processes in the world that are represented—and that this might account for the utility of mathematical concepts.

I recall he more or less asked, "Is it really so surprising that certain patterns of neurophysiological activity, that are in fact what mathematical concepts are at the level of the human brain, can be mapped onto the world?" Perhaps there's some kind of sameness of structure or homology there. Is that a step toward resolving this mystery for you, or do you think it exceeds that?

MT: That's an interesting argument—the argument that our brain adapts to the world and therefore has the world model inside it.

SH: Our brains are just a part of the world. There are processes in the world that have a similar structure to brain processes that constitute our mathematical intuitions, and this allows for some kind of mapping.

MT: Yeah. I agree with the first part of the argument and disagree with the second part. I agree that it's natural that there will be things in the brain that are very similar to what is happening in the world, precisely because the brain has evolved to have a good world model. But I disagree that this fully answers the whole question.

Because the claim that he made there, that brain processes of certain kinds are effectively what mathematics is—most mathematicians I know would violently disagree that math has something to do with brain processes at all. They think of math rather as structures which have nothing to do with the brain.

SH: Let's pull the brakes there. Clearly, your experience of doing math, your grasp of a mathematical concept or not, the moment something makes sense or you persist in your confusion, your memory of the multiplication table, your ability to do basic algebra and everything beyond—all of that, in every instance of its being realized, is realized as a state of your brain. You're not disputing that?

MT: Of course. Absolutely. I'm just quibbling about what mathematics is. What's your definition of mathematics? I think it's interesting to take a step back and ask,
"What do mathematicians today generally define math as?" Because if you ask people on the street—like my mom, for example—they will often view math as just a bag of tricks for manipulating numbers or maybe as a sadistic form of torture invented by schoolteachers to ruin our self-confidence.

Whereas mathematicians talk about mathematical structures and studying their properties. I have a colleague here at MIT, for example, who has spent 10 years of his life studying this mathematical structure called E8. Never mind what it is, exactly. But he has a poster of it on the wall of his office. David Vogan. If I went and suggested to him that that thing on his wall is just something he made up somehow, that he invented it, he would be very offended. He feels he discovered it. That it was out there, and he discovered that it was out there, and mapped out its properties in exactly the same way that we discovered, rather than invented, the planet Neptune and then went out to study its properties. Similarly, look at something more familiar than E8, the counting numbers—one, two, three, four, five, etc. The fact that two plus two is four, and four plus two is six. Most mathematicians would argue that this structure, this mathematical structure that we call the numbers, is not a structure that we invented or invented properties of, but rather that we discovered the properties of. This has been discovered multiple times independently in different cultures. In each culture, people invented, rather than discovered, a different language for describing it.

You know, in English you say one, two, three, four, five. In Swedish, the language I grew up with, you say *ett, två, tre, fyra, fem.* But if you use a Swedish-English dictionary and translate between the two, you see that these are two equivalent descriptions of exactly the same structure. Similarly, we invent symbols. What symbol you use to write the numbers two and three are actually different in the U.S. versus in India today, or in the Roman Empire. Again, once you have your dictionaries there, you see that there's still only one structure that we discover, and then we invent languages.

SH: Yes. Of course.

MT: To just drive this home with one better example: Plato was really fascinated by these very regular geometric shapes that now bear his name: Platonic solids. He discovered that there were five of them: the cube, the octahedron, the tetrahedron, the icosahedron, and the dodecahedron. He chose to invent the name "dodecahedron." And he could have called it the "schmodecahedron," or something else. But that was his prerogative—to invent names, the language for describing them. But he was *not* free to just invent a sixth Platonic solid.

SH: No doubt.

MT: Because it just doesn't exist. So it's in that sense that Plato felt that those exist out there and are discovered rather than invented. Does that make sense?

SH: Yes. I certainly agree with that. I don't think you actually have to deny that mathematics is a landscape of possible discovery that exceeds our current understanding—and, in fact, will always exceed it. What is the highest prime number above the current one we know? Well, clearly, there's an answer to that question.

MT: You mean the lowest prime number above all the ones we know?

SH: Sorry, yes, I meant the *next* prime above the highest one we know.

MT: Yeah. Yeah.

SH: That number will be discovered rather than invented. Or, to invent it would be to invent it within the perfect constraints of its being, in fact, the next prime number. So, it's not wrong to call that a matter of pure discovery, more or less analogous, as you said, to finding Neptune when you didn't know it existed, or going to the continent of Africa. Africa is there whether you've been there or not.

MT: Right.

SH: So, I agree with that. But it still seems true to say that every instance of these operations being performed, every instance of mathematical insight, every prime number being thought about or located—every one of those moments is a moment of a brain doing its mathematical work.

MT: Right. Or a computer, sometimes.

SH: Yes.

MT: Because we have an increasingly large number of proofs now done by machines. And discoveries also, sometimes.

SH: Yes. We're still talking about physical systems that can play this game of discovery in this mathematical landscape that exists, in some sense, whether or not there is anyone to do the discovering.

MT: Right.

SH: But the fundamental mystery is, why should mathematics be so useful for describing the physical world and for making predictions about blank spaces on the map?

MT: Exactly.

SH: Again—and I'm kind of stumbling through this conversation, because I'm not a mathematician, nor am I a philosopher of mathematics. I'm sort of shooting from the hip here, but I want to get a sense of whether this could remove some of the mystery, if there are certain physical processes in brains and computers and other intelligent systems, wherever they are, that can mirror this landscape of potential discovery. Does that remove what otherwise seems spooky and Platonic—the mystery that abstract, idealized concepts fit the physical universe?

MT: That's a great question. The answer you're going to get to that question will depend dramatically on who you ask. There are very, very smart and respectable people who come down all across a very broad spectrum of views on this. In my book, I chose not to say "This is how it is" but, rather, to explore the whole spectrum of opinions. Some people will say, if you ask them about this mystery, "There is no mystery. Math is sometimes useful in nature, sometimes it's not. That's it. There's nothing too mysterious about it. Go away."

If you go a little bit more towards the Platonic side, you'll find a lot of people saying things like "Well, it seems like a lot of things in our universe are very actively approximated by math. And that's great. But they're still not perfectly described by math." And then you have some very, very optimistic physicists, like Einstein and a lot of string theorists, who think that there actually is some math that we maybe haven't discovered yet. That doesn't just approximate our physical world but describes it exactly and is the perfect description.

And then, finally, the most extreme position on the other side—which I explore at length in the book, and that's the one I'm personally guessing on—is that not only is our world described by mathematics, but it *is* mathematics, in the sense that the two are really the same. You talked about how, in the physical world, we discover new entities and then we invent language to describe them. Similarly, in mathematics, we discover new entities, like new prime numbers, Platonic solids, and we invent names for them.

Maybe this mathematical reality and the physical reality are actually one and the same. When you first hear that, it sounds completely Looney Toons, of course. It's equivalent to saying that the physical world doesn't just have some mathematical properties, it *only* has mathematical properties.

That sounds really dumb if you look at your wife and your child, or whatever, and think, "This doesn't look like a bunch of numbers." But to me, as a physicist, when I look at them… When I met Annaka, your wife, for the first time, of course she has all these properties that don't strike me as mathematical.

SH: Don't tell me you were noticing my wife's mathematical properties.

MT: (laughing) But at the same time, as a physicist, I couldn't help noticing that your wife was made entirely out of quarks and electrons. What properties does an electron actually have? Well, it has the property -1, 1/2, 1, and so on. And we've made up nerdy names for these properties, we physicists, such as electric charge, spin, and electron number. But the electron doesn't care what language we invent to describe these numbers. The properties are just these numbers, just mathematical properties. And for Annaka's quarks, same deal. Also, the only properties they have are also numbers, except different numbers from the electrons'. So the only difference between a quark and an electron is what numbers they have as their property. And if you take seriously that everything in both your wife and the world is made of these elementary particles that have only mathematical properties, then you can ask, "Then what about the space itself that these particles are in? What properties does space have?" Well, it has the property three, for starters. You know, the number of dimensions. Which, again, is just a number.

Einstein discovered that it also has some properties called "curvature" and "topology," but they're mathematical too. If both space itself and all the stuff in space have only mathematical properties, then it starts to sound a little bit less ridiculous, the idea that maybe everything is completely mathematical and we're actually a part of this enormous mathematical object.

SH: I don't want to spend too much more time here, because there are many other things I want to get into in your book. But this is a fascinating area for me—again, unfortunately, one that I feel unequipped to have strong opinions about. How is what you just said different from saying that every *description* of reality we arrive at—everything one can say about quarks, or space, or anything—is just a matter of math and values? Couldn't we also say that it's just a matter of English sentences, or sentences spoken in human language. In other words, could the question of why mathematics is so good at representing reality be a little like the question of why language is so good for speaking in, or so good for capturing our beliefs? Is there a dis-analogy there that can save us?

MT: The language we invent to describe mathematics—the symbols for the numbers and for plus and multiplication and so on—is of course a language too. Languages generally are useful, yes. But there's a big difference. Human languages are notoriously vague. And that's why the radio and the planet Neptune and the Higgs boson were not discovered by people just sitting around blah blah blah-ing in English, but with the judicious use of the language of mathematics.

And all three of these objects were discovered because someone sat down with a pencil and paper, did a bunch of math, and made a prediction. If you look over there at that time, you'll find Neptune, a new planet. If you build this gadget, you'll be able to send radio waves. If you build this Large Hadron Collider, you'll find a new particle. There's real power in there.

Before we leave this math topic, I want to end on an emotional note. Some people don't like this idea because they think it sounds counterintuitive. We already laid that to rest at the beginning of our conversation. Other people don't like it because they feel it sort of insults their ego. They don't want to think of themselves as a mathematical entity or whatnot.

But I actually think this is a very optimistic idea if it's true. Because if it's wrong, this idea that nature is completely mathematical, that means that the quest of physics, which has exploited the discovery of mathematical patterns to invent new technologies, is going to end eventually—that physics is doomed. One day we'll hit this roadblock when we've found all the mathematical patterns there were to find. We won't ever get any more clues from nature. And then we can't go any further with our understanding or technology. Whereas if it's all math, there is no such roadblock, and the ability of life in the future to progress is really only limited by our own imagination. To me, that's the optimistic view.

SH: Is there any connection between this claim that it's all math at bottom and the claim that it's all information? I'm now hearing echoes of John Wheeler, who talked about "it from bit"—this concept that at some level, the universe is a computation. Is there a connection between these two discussions, or are they distinct?

MT: Yeah, there probably is. John Wheeler is one of my great heroes. I had the great fortune to spend a lot of time with him when I was a postdoc at Princeton, and he really inspired me. My hunch is that we will one day in the future come to understand more deeply what information really is and its role in physics—and also come to understand more deeply the role of computation and quantum computation in the universe. We will one day come to realize, maybe, that mathematics, computation, and information are just three different ways of looking at the same thing. We're not there yet. But that would be my guess.

SH: Are we there on the topic of entropy? Is there a relationship between entropy in terms of energy and entropy in terms of information? Is there a unified concept there, or just an analogy bridging those two discussions?

MT: I think that's fairly well understood, even though there are still some controversies brewing. This is a very active topic of research. In fact, you mentioned that you and I met at a conference that I was involved in organizing. The previous conference I organized was called "The Physics of Information." We brought together physicists, computer science people, neuroscientists, and philosophers, and had a huge amount of fun discussing exactly these questions.

So I think there's a lot more to come. To me, these ideas, the most far out and speculative ideas I explore in the book about the role of math, are not to be viewed as the final answer to end all research but, rather, simply as a great way to generate new, cool, practical applications of things. They're a roadmap to finding new problems. You hinted at some of them here. I think there are a lot of fascinating relationships between information and computation and math and the world that we haven't discovered yet. It would probably have a lot to do with how consciousness works as well, is my guess. I think we have a lot of cool science to look forward to.

SH: Consciousness is really the center of my interest. But we may not get there, because I now want to get into the multiverse, which is probably the strangest concept in science now. It's something I thought I understood before picking up your book, and then I discovered there were three more flavors of multiverse than I realized existed.

I want to talk about the multiverse, but first let's start with the universe, because this is a term around which there's some confusion. Let's just get our bearings. What do we mean, or what *should* we mean, by the term "universe"? And I want to start with your level-one multiverse. So, if it's possible, give us a brief description of the concept of inflation that gets us there.

MT: Sure. So, first of all, what is our universe, before we start talking about others? Many people tacitly assume that "universe" is a synonym for everything that exists. If so, by definition, there can't be anything more. Talk of parallel universes would just be silly, right? But that is not in fact what people in cosmology mean when they say "universe."

When they say, "our universe," they mean the spherical region of space from which light has had time to reach us so far during the 13.8 billion years since our Big Bang. In other words, everything we could possibly see, even with unlimited funding for telescopes. If that's our universe, we can reasonably ask, "Well, is there more space beyond that, from which light has not yet reached us but might reach us tomorrow, or in a billion years?"

And if there is—if space goes on far beyond this, if it's infinite or just vastly larger than the space we can see—then all these other regions, which are as big as our universe if they also have galaxies and planets in them and so on, it would be kind of arrogant to not call them universes as well. Because the people who live there will call that their universe.

Inflation is very linked to this, because it's the best theory we have for what created our Big Bang and what made our space the way it is, so vast and so expanding. It actually predicts, generically, that space is not just really big but vast, and in most cases, actually infinite. Which would mean that if inflation actually happened, what we call our universe is really just a small part of a much bigger space.

So, in other words, space is much bigger than the part of space that we call our universe. This is something I don't actually think is particularly weird, once we get the terminology straight. Because it's just history all over again, right? We humans have been the masters of underestimation.

We've had the overinflated ego, where we want to put ourselves in the center and assume that everything we know about is everything that exists. And we've been proven wrong again and again and again, discovering that everything we thought existed is just a small part of a much grander structure.

A planet, a solar system, a galaxy, a galaxy cluster, our universe, and maybe also a hierarchy now of parallel universes—it would just continue the same trend. And for somebody to object on some sort of philosophical grounds that things can't exist if they're outside our universe, if we can't see them, seems very arrogant. Much like an ostrich with its head in the sand, saying, "If I can't see it, it can't exist."

SH: Right. But things begin to get very weird, given this fact that inflation, as you said, is the best current picture of how things got started. Given that inflation predicts a universe of infinite extent—infinite space, infinite matter—you have, therefore, a universe in which everything that is possible is, in fact, *actual*. Everything happens. Everything happens, in fact, an infinite number of times. Which is to say that you and I have this podcast an infinite number of times and an infinite number of different ways.

We're still talking about the level-one multiverse here. So we're just saying that if you could travel far enough, fast enough, you would arrive on some planet disconcertingly like Earth, where you and I are having a virtually identical podcast but for a single change in term, or I might just decide to shave off my eyebrows in the middle of this conversation.

MT: Exactly. Or I switch to talk in French.

SH: This is…well, let's stop there. Is that, in fact, what you think a majority of cosmologists believe?

MT: This is a great question. First, that's a great illustration of one of the cool things in science, where you start with some pretty innocent assumptions—namely, here, that space just goes on forever, like most of us thought as kids. And, moreover, that things started out a little bit randomly everywhere.

And you get this totally shocking conclusion. When I ask my colleagues, the vast majority of them will put their money on that some form of inflation happened, and our space is actually much bigger than our universe. Whether it's actually infinite or just really huge starts getting a little bit more controversial.

We also don't know for sure, of course, whether inflation actually happened. But this is the simplest version of the theory: Space simply goes on forever. It's an infinite space. Much like Euclid's space, or the one we thought about as kids. In the book, I call this the level-one multiverse. But you can use "space" as a synonym for it.

Just to drill down a little bit more on where the craziness comes in, if you look at how our universe got this way, and how our podcast came about, it's because we had about ten to the 78th power quarks and electrons that started out in a particular way, somewhat random early on, after inflation. Which led to the formation of our solar system and our planet. And our parents met, and so on, and we met, and then this happened, right?

If you'd started the quarks out a little bit differently, things would have unfolded differently. You can actually count up how many different ways you can arrange the quarks and electrons in our universe. It turns out it's only about a googolplex different ways. A googolplex is one with a google zeroes, and a google is one with 100 zeroes. So it's a huge number, but it's finite.

So if you have an infinite number of other regions equally big, and you roll the dice again in all of them, you can calculate that if you go about a googolplex meters away, you will indeed end up with just what you described: a universe that's extremely similar to this one except that one minute ago you all of a sudden decided to start speaking Hungarian instead. It's a very mind-boggling idea. We don't know for a fact that it's like this. But this is sort of the vanilla-flavored cosmological model, the one that is most popular today.

SH: Well, I think the weak link in this chain of reasoning, or the place where a skeptical person can get off the train, is in the assumption that inflation implies an infinite universe rather than just a very large one. So it seems like you could pull the brakes there. But unfortunately, this concept of a multiverse, judging from your discussion of it in your book (and this is what I didn't understand before I picked up your book), seems overdetermined. There are other ways of arriving at this multiverse concept, which we'll get to. Scientifically speaking, there are many reasons to believe in a functionally infinite number of copies of ourselves living out lives, for all intents and purposes, that are exactly similar or differing to every possible degree, right? So it's true that everything that can happen does happen under this rubric?

MT: That's right. To distinguish the bit where we know and where we don't know for sure: The part we don't know for sure is that space is infinite—that there's an infinite number of anything. For people who feel really bothered by these implications and want to get rid of the infinity, I have a whole section in the book where I attack infinity and list all the ways in which you can get rid of it.

So there are a lot of interesting opportunities there, and we're going to know more, I think, in the next five or ten years. However, what seems pretty much inescapable at this point is that the full reality is at least much larger than what we can see. There's just no way that space ends exactly at the edge of our universe.

In fact, if you had made that claim one minute ago, I could falsify it now by looking with a telescope. Because you can see light that's traveled from one minute farther away. And that's pretty far. That's an eighth of the way to the sun already. So we should probably get used to the idea that we live in a much grander reality than we thought we did. And I think that's a good thing.

SH: I don't think people's intuitions recoil at the very, very large. I even think people are prepared to embrace the infinite and the eternal, in some sense. Even though we could debate whether thinking about a beginning is actually more understandable than thinking about an eternal universe, given how squirrelly the beginning begins to look.

But I think what really will blow the mind of anyone who thinks about it long enough is what is implied by the sheer concept of infinity—that everything that is possible is, in fact, actual. On some level, everything is true. Let's spell out why this should be disturbing and why it may be, at least at first glance, a real embarrassment to science. Because science prides itself on being parsimonious.

MT: Right.

SH: Seemingly, this is not only un-parsimonious, this is the least parsimonious idea imaginable.

MT: I disagree, actually.

SH: Well, I want to get there, but I don't want to leave our listeners confused. Some might not even know what I mean by "parsimonious," because in common parlance, that just means "stingy" or "not wasteful." In science, we talk about a theory, or the style of thinking that is scientific, as being parsimonious in not being theoretically wasteful.

Occam's razor, which many people have heard of, admonishes us not to multiply quantities without reason. So if the concept of gravity can explain why it's difficult to lift massive objects, then just use gravity and don't posit the existence of invisible elves that are also holding them down. That's why we're biased toward simple explanations over what seem like needlessly complex ones.

MT: Exactly.

SH: On its face, this implication of infinity seems incredibly un-parsimonious. Because again, we're saying that essentially everything that's possible is true. I guess in one sense this can seem parsimonious when you think of having to invite people to a party. You're coming up with a guest list for a party, and you're faced with many hard decisions, or you could say…

MT: You could invite everybody.

SH: Yes, just invite everybody. Or even worse than that, you could say, "Well, let's invite everyone on earth and then just call whatever they're already doing the party." That, in fact, is simpler than coming up with a guest list that excludes people. So tell me why this is not embarrassing with respect to constraints like Occam's razor.

MT: With pleasure. I'm a big fan of Occam's razor. You'd be pleased to know that I actually have a framed equation of gravity here in my office. If I were to add some elves to that, it would make the equation framed here much uglier and more complicated.

To me, Occam's razor means that you don't want to add wasteful things to your theory. You want to keep it as simple as you can. So let's drill down and ask, what is it that we feel is so wasteful in this inflationary universe? Is it that we're worried about wasting space? Hardly. Because even Newton's theory of physics had an infinite space, right? Space was just the space of Euclid. It goes on forever in all directions. Was it that we were worried about wasting atoms? No. Because again, that's what a lot of people thought we had earlier also.

But somehow, I think, people feel that it's wasteful in terms of information. It just sounds so complicated that you have to describe all possible ways in which you and I could have this conversation and so on. In physics, what we really, really value the simplicity of is not the solutions to the equations but the equations themselves. The fact that we can write down equations that can describe everything around us in the world on a blackboard. That's the parsimony.

The theory of inflation, like the theory of general relativity, is extremely simple and parsimonious. That's why it's become so popular. Because you get much more out of it than you put into it. You put in these very simple equations, you can predict all that stuff. If you go a little further—you add the standard model of particle physics and so on—it turns out that with just those equations and a little cheat sheet with 32 numbers, we can calculate hundreds of thousands of numbers.

Every single number we've ever measured in the physics labs around the world—that's parsimony, in that the equations are simple. The math is simple. Never mind that the solutions are complicated. Think of Niagara Falls. The equations that describe the water flow there are called the Navier-Stokes equations, and they're simple enough that you could put them on your T-shirt.

But look at the solution. It's so complicated, with all the spray and all the water droplets, and the turbulence. Yet we feel that this is a perfectly beautiful explanation of what's happening, these equations. Because the equations are simple.

SH: Right. Well, let's press on into the multiverse. This will push people's intuitions in the direction of feeling that at the very least, we're trying to have our cake and eat it too on this question of parsimony. Take us to the level-two multiverse and perhaps say why this is relevant to the question of fine-tuning, which, many people will recall, is connected to this idea that many religious people have of why religion—the idea of a creator God in particular—makes sense, given the apparent fine-tuning of our universe.

MT: With pleasure. First, though, let me say one more thing about the level-one multiverse here so that listeners don't worry too much about what you said about everything happening somewhere. Someone might worry that they're an ax murderer in a parallel universe.

SH: Oh, they are.

MT: But I want to put them at ease by saying that if you start traveling through space, these vast distances, and you find all these other planets where other things are happening, on the vast majority of them, their alter egos are not ax murderers. They're doing very, very reasonable things. So, worried people, they are really the rare flukes.

SH: But wait a minute… The rare flukes are not only true, there are an infinite number of them. It's just, proportionally speaking, it's a lesser infinity than the other infinities of apparently benign, lawfully behaving selves.

MT: Yeah. But it's kind of like when I'm on an airplane, if I worry that it's going to crash. I know that in the multiverse it *is* going to crash and it's *not* going to crash. But that doesn't traumatize me particularly, because I know that the fraction of all the parallel universes where it's going to crash is much less than one in a million anyway.

SH: Right.

MT: Only a very tiny fraction of the Maxes get wiped out, and I don't worry about it. It's exactly in that way that people shouldn't worry too much about being ax murderers somewhere else.

SH: Yes, proportionality is what we care about.

MT: Exactly. So again, the level-two multiverse, you can simply call it "space," if you want. Inflation is actually not only able to make an infinite space, but also able to make fit within it an infinite number of regions, each of which seems infinite to whoever lives inside it, through some very weird properties of Einstein's gravity theory that I talk about in the book.

What's interesting about this is that when you ask how diverse space is, you might think that in some places our podcast goes here, and in other places we talk about other things, but at least the laws of physics are the same everywhere. You might think that if people learn different things in history class—if the Sam Harris somewhere else learns different things in his history class, because the quarks started out differently there and history played out differently—at least they're going to learn the same thing in physics class.

But the level-two multiverse changes that also. Because it turns out that a lot of things that we thought were fundamental laws of physics—that were true everywhere in space—were actually not, it seems. I like to think about it as if I were a fish swimming around in the ocean, I would think that it's a law of physics: The water is something you can swim through. Because that's the only kind of water I know, and it seems to be that way everywhere I look. But if I were a really smart fish, I could solve equations—discover the equations for water, solve them, and see that there are actually three solutions, not one. There's the water solution and then also ice and steam.

Equivalently, there are a lot of hints now in physics that what we call empty space is also like that: It can freeze and melt and come in many different variants. And the thing is, inflation is so violent that if space actually can be in many different forms, it will create each of those kinds of space, and an infinite amount of it at that.

So if you go really, really far away, you might find yourself in a part of space where there are not actually six kinds of quarks, like there are here, but maybe 10 kinds of quarks. So the level-two multiverse is very, very diverse. Also, a lot of things that we learn in school are fundamental parameters of physics. For example, the number 1,836 seems kind of hardwired into our world—the proton is 1,836 times heavier than the electron.

Why is that? Well, string theory suggests that actually that's one of those things that also changes depending on what kind of space you have. So it might be 2015 somewhere else, and 666 somewhere else. This explains the fine-tuning problem you've mentioned. Because we've discovered, as I mentioned earlier, that there are at least 32 numbers—pure numbers, with no units or anything, that we've measured—that we can use to calculate everything else. And we wonder a lot about where they came from.

SH: So these are the constants of nature? Could you list a few of them to give people a sense?

MT: Yeah. So 1,836 is one of them—how much heavier the proton is than the neutron. You can transform them in different ways. Another one, which is often talked about these days, is the density of dark energy, which makes up about 70% of all the stuff in our universe. And it turns out, if you think of each of these parameters as written on a knob that you could twist, you shouldn't touch the knobs. That's my advice. Because if you tweaked most of them, life as we know it would be completely destroyed. The sun would explode, or something else very bad would happen. That's the fine-tuning you mentioned—it seems like many of these parameters have been dialed in at exactly the right value needed to support life.

SH: And for some of them, the fine-tuning is incredibly fine. We're talking about, what, ten decimal places or beyond, right?

MT: Yeah. Even for something as basic as how strong the electric force is, if you changed it about a percent one direction or another, then you wouldn't have enough oxygen or enough carbon any more to have life on earth. And the most fine-tuned one of all is this dark energy I mentioned, which is fine-tuned to over 100 decimal places. It's ridiculous.

SH: Religious people are getting very excited here…

MT: Some are, but not all. The level-two multiverse gives an alternative explanation. Because if it's actually the case that this number is just dialed in randomly in different huge swaths of space, then since you can calculate that you're only going to form galaxies in those places where the dark energy is just right—where you're in this bio-friendly Goldilocks zone of dark energy—then of course this question of how much dark energy there is, is only going to be asked in those places where there's life.

So we shouldn't be surprised at all that we actually find dark energy to be in that life-friendly range. You know, basically, we hate unexplained surprises in science. When we get an unexplained surprise, we usually say that means our theory has been ruled out. Right? And what this does is actually give an explanation of why we're measuring these values. The famous physicist Steven Weinberg even used this argument to predict the amount of dark energy here before they actually discovered dark energy. And it turned out his prediction was really quite good.

So that's one of the feathers in the hat of the level-two multiverse. We don't yet know, of course, whether it exists, but it's kind of hard to get rid of. If string theory turns out to be wrong, the competitor Loop quantum gravity also seems to have multiple solutions for space.

And it's a pretty general property of math that if you have some complicated equations, they have many solutions. Inflation has this amazing property of being this creative force that transforms potential existence into actual existence. So any kind of space that can exist, inflation will create a huge swath of it and thereby be the great enabler, basically, of possibilities.

SH: So, to connect this level-two multiverse with the level one multiverse: In the level one, we were talking about the universe as we know it extending infinitely or almost infinitely beyond the horizon of what we can make out. So we were really just talking about more space and more matter.

MT: That's right.

SH: Which, if it's infinite, suggests that everything that can happen within the laws of physics does happen. With level two, we're talking about inflation creating an infinite number of bubble universes, wherein the laws of physics themselves vary in every conceivable way and we…

MT: Well, let me just interject there so it doesn't sound too weird: Instead of talking about "bubble universes," we can just keep saying "space." Because there's still only one space.

SH: But it's not space in a straightforward sense, is it?

MT: No, it is, actually. But the reason we can never get to another part of the level-two multiverse is because in order to go there, you would have to go through a region of space that's still inflating and stretching out. So, if you have your kids in the backseat, asking "Are we there yet?" you say, "Oh, yeah, we'll be there in one hour." And then a little bit later, they ask, "Are we there yet?" And you say, "Yeah, we'll be there in two hours."

SH: Right.

MT: So inflation can actually create this funny situation where you have many, even infinite, regions of space that are still fitting into one single piece of space. That's one clarification. It's still just this one space, but messy. And the second thing is, it's not that the actual laws of physics are different. It's just that things that we thought were laws of physics turned out to actually be different *solutions* to the laws of physics.

SH: Exactly.

MT: Ice is not a different law of physics from liquid water or steam. They turn out to be three different solutions to the equations for water. This is a cool trend in science, where… for example, Kepler, very smart guy who figured out that the planets go around an ellipse. He tried to predict, from first principles, why the orbits of the planets in our solar system were the size that they were and came up with a really beautiful theory that—it's something like you put a cube inside a dodecahedron, inside an icosahedron, etc. It's supposed to match up with Mercury and Venus and Mars and Earth and Jupiter, etc.

Now, people would just laugh at that. Because it doesn't make sense that you should be able to predict exactly the orbit of Earth and the orbit of Mars, because there are many solar systems where the answers are different. Right? So how could you possibly? It's not like the size of Earth's orbit is something that's in the laws of physics.

It just came about sort of randomly from the way our solar system was born. And what the level-two multiverse does is similarly downgrade a lot of other stuff that people thought were fundamental laws of physics, like how many quarks there are. Actually it says that, too, is a historical accident that had to do with the way this region of space got created.

SH: Right. It's actually shuttling some of the subject matter of physics into the subject matter of history, albeit history of a very erudite kind.

MT: That's exactly right. And that's something Occam would like, because it makes physics itself simpler, and it makes history more complicated.

SH: That's a fascinating idea. It closes the door to this otherwise embarrassing problem of fine-tuning—the question of how it is that we find ourselves in a universe that seems perfectly tuned to support life and intelligent life and beings exactly like ourselves in a position to wonder about these things? There have been other efforts to close that door just with what's been called the anthropic principle, which you stated earlier:
The only place we can find ourselves is a place that's compatible with our existence. That shouldn't be surprising. And yet it has seemed surprising that, essentially, we should exist at all—that the universe could have been an infinite number of ways and it just happened to be this way.

Well, according to the level-two multiverse, the universe *is* essentially an infinite number of ways, and there are an infinite number of regions that are not compatible with life. It's a Darwinian principle of universe emergence—that the only places you can find yourself are the places you can find yourself, and every place that is possible exists.

MT: Yes. I don't like the use of the term "anthropic principle" for these sort of things, because the word "principle" makes it sound like it's somehow optional. I mean, it's just the correct use of logic, which, of course, is not optional. Are you really, really surprised that out of all the eight planets in our solar system, we're living on Earth rather than Venus, where it's 900 Fahrenheit right now? Or on Jupiter, where there's no surface to stand on? You're probably not very surprised. I wouldn't call that some deep principle. It's just common sense that the vast majority of our solar system is not very friendly for our kind of life. And the vast majority of space is horrible for our kind of life. Therefore, we shouldn't be surprised that we're living in a very special part of space that we can see.

SH: I agree. It's a kind of pseudo mystery based on a needlessly post hoc look at probability. I could ask the same question with respect to the arrangement of objects on my desk at this moment. Objects are strewn everywhere. I have a very messy desk. What are the chances that the pen would be exactly where it is in this moment?

MT: Exactly. Or what are the chances that your mother's mother's mother met your mother's mother's father, given the way things were 50 years before that?

SH: Right.

MT: Look at all the coincidences that seem to happen. But I'm sure you don't lose a lot of sleep over that one.

SH: Another way of closing the door to this mystery of fine-tuning, which doesn't entail a level-two multiverse, is the idea that we could be in a simulation. I don't know if this argument originates with your friend Nick Bostrom. I know other people have arrived at this independently.

MT: The argument is older. But Nick Bostrom made a very detailed argument for why he thinks it's actually likely that we live in a simulation.

SH: His argument, in brief, is that if we imagine ourselves in the distant future, with beings like ourselves that have made vast gains in their ability to produce computers, it stands to reason that they will simulate universes and beings very much like ourselves on those computers—assuming that such a thing is possible. And there's really no reason to think it isn't. And then, just by dint of numbers, you would expect simulations to vastly outnumber real universes.

And so you would expect that we are now in a simulation rather than in a real universe. That argument sort of stands on its own, unrelated to this issue of fine-tuning or the multiverse. But if taken seriously, the prospect of being in a simulation answers this fine-tuning argument as well, correct?

MT: It's a fascinating question. I give a detailed argument in the book for why I think we're *not* living in a simulation. I won't get into it now, in the interest of time. But just to get a sense for what might be fishy here, suppose you buy the simulation argument. And you say, "Okay, we are living in a simulation. This is not the actual reality. There's some sort of basement reality where computers are simulating us with different laws of physics."

Suppose you buy that. Then you can make the argument again that you should be a simulation within the simulation, for exactly the same reasons. And you can repeat the argument ad absurdum and then decide that you're a simulation within a simulation within a simulation, almost forever. Something seems fishy there.

As I explain in more detail in the book, the fundamental flaw in the argument, I feel, is that if we *are* in a simulation, there's no reason to believe that the laws of physics that the simulating computers are obeying have anything to do with the laws of physics that we see around us in this world. Because this isn't the real reality where the simulation is happening.

And the simulation argument kind of conflicts these two. Finally, though, in case you're still worried about living in the simulation, I'll give you some advice. Live a really interesting life and do interesting things so that whoever's running it doesn't get bored and shut you down.

SH: Well, my concern, given my record of being critical of religion, is that if the Mormons are running the simulation at some point in the future, I could be living in a simulation where Mormonism, or any other religious conception of reality, is precisely true. So, see you in hell, Max.

MT: That's your vision of purgatory, is it?

SH: This might be a good bridge to a final topic. I'm mindful of your time here, and we're not doing anything like justice to the contents of your book. We'll skip over the other ways in which we can arrive at a conception of a multiverse—the quantum mechanical issues addressed by Hugh Everett, and all of that, are fascinating—it's just another route into infinite copies of ourselves having infinite versions of this podcast, and no doubt in some of those podcasts we treat these topics at much greater length. But I think this is a good bridge to AI, which is why you and I met at the conference you organized through your institute. I came into that conference as an utter novice on this topic. I had more or less ignored AI, having accepted the rumors that little progress had been made—all the promises had been overblown, and there was not much to worry about—it was just a dead end, scientifically.

And then I heard our mutual friend, Elon Musk, and other people like Stephen Hawking, worrying out loud about the prospect of AI. And whether it's five years or 50 years, in a time frame that any rational person—certainly any rational person who has kids—could worry about, the fear is that we will make huge gains that could well destroy us if we don't anticipate the ways in which machines more intelligent than ourselves could fail to converge with our interests and or be controllable by us.

I've mentioned this on the podcast a few times, and I've recommended Nick Bostrom's book on this topic, *Superintelligence,* which is really a great summary of the problem. You and I both answered the Edge Question, my response to which is also on my blog. The Edge Question was on this topic right after the conference in San Juan that you organized.

I noticed that there are many smart people, many of whom should be very close to the data here, who are deeply skeptical that there's anything to worry about. Friends and colleagues of mine, and perhaps yours, like Steven Pinker and Lawrence Krauss, take a very different line here and have more or less said that the safety concerns about AI are totally overblown—we'll just get into the end zone and figure it out. I mean, they're basically treating it like the Y2K scare. I'm wondering what you think about that and what accounts for this difference of opinion.

MT: This is fascinating. I've noticed this, too. First of all, these questions are so unfamiliar that a lot of very smart people actually get confused about them. It's also interesting to be clear on the fact that people who say, "Don't worry" very often disagree with one another. You have, for example, one camp who say, "Let's not worry because we're never going to get machines smarter than people," or at least not for hundreds of years.

This camp includes a lot of famous businesspeople and a lot of great people in the AI field also. You had Andrew Ng, for example, saying that worrying about AI becoming smarter than people and causing problems is like worrying about overpopulation on Mars. He's a good ambassador for that camp. And you have to respect that. It might very well be that we will not get anything like human-level AI for hundreds of years.

Then you have another group of very smart people who say, "Don't worry" for sort of the opposite reason. They say, "Let's not worry. We are convinced that we are going to get human-level AI, probably in our lifetime with good odds, but it's going to be fine." I call them the digital utopians. And there's a fine tradition in this also.

You have a lot of beautiful books by people like Hans Moravec, Ray Kurzweil. And a lot of leading people in the AI field fall into that camp. They think that AI will succeed. That's why they're working on it so hard right now, and they're convinced that it's not going to go wrong.

So, for starters, I would love to have a debate between these two groups of people, who don't worry about why they differ so much in their timelines. My own attitude about this is, I agree, we certainly don't know for sure that we'll get human-level AI, or that if we do, it's going to be a great problem. But we also don't know for sure that it won't happen.

And as long as we are not sure that it won't be a disaster in our lifetime, it's good strategy to pay some attention to it now. Even if you're figuring your house probably won't burn down, it's still good to have a fire extinguisher and not leave the candles burning when you go to bed. You know, take some precautions.

That was very much the spirit of this conference: Look at concrete things we can do now to increase the chances of things going well. And finally, we have to stress that as opposed to other things you could worry about, like nuclear war or some new horrible virus or whatever, this question of AI is not just negative. It's also something that has a huge potential upside.

We have so many terrible problems in the world that we're failing to solve because we don't understand things well enough. If we can amplify our intelligence with artificial intelligence, it will give us great power to do things better for life in the future. But like any powerful technology, it can be used for good. It can also be used, of course, to screw up.

But when we've invented less-powerful tech in the past, like when we invented fire, we learned from our mistakes. And then we invented the fire extinguisher, and things were more or less fine, right? But with more-powerful tech like nuclear weapons, synthetic biology, future super-advanced AI, we don't want to learn from our mistakes. We really want to get it right the first time.

SH: Yes.

MT: Because it might be the only one we have.

SH: In my view, and in the views of many others, that's what makes this AI issue unique. Because we're ultimately talking about autonomous systems that will exceed us in intelligence. And as you say, the temptation to turn these systems loose on the other problems that we confront is going to be exquisite. Of course we want something that can help us cure Alzheimer's—or cure Alzheimer's on its own. We want something that can stabilize the global economy, or give us a perfect climate science, etc.

There's nothing better than intelligence. And to have more of it would seem an intrinsic good. Except, if you imagine getting what the mathematician I.J. Good described as an "intelligence explosion," where this thing could get away from us, and we would not be able to say, "Oh, no, sorry, that's not what we meant. Here, let's modify your code."

MT: Exactly.

SH: But many smart people doubt that any sort of intelligence explosion is possible. That's the sense I'm getting. They view it very much like other things—like fire or nuclear weapons—where all technology is powerful and you don't want it to fall into the wrong hands, where people use it maliciously or stupidly. But we understand that. And they think it doesn't really go beyond that.

People trivialize this by saying that there’s no reason to think that computers are going to become malicious—spawning armies of Terminator robots because they decide they want to kill us. But that’s not the fear. The fear isn’t that they’ll spontaneously become malevolent, it’s that we could fail to anticipate some way in which their behavior could diverge, however subtly but fatally, from our own interests. And this thing could then get away from us in a way that we could no longer correct for.

MT: Exactly. We should not fear malevolence. We should fear competence. Because what is intelligence to an AI researcher? It’s simply being really good at accomplishing your goals, whatever they are. A chess computer is considered very intelligent if it’s really good at winning in chess. There’s another game, called Losing Chess, which has the opposite goal, where you try to lose. And their computer is considered intelligent if it loses the games better than any others.

So the goals have nothing, really, to do with how competent it is. And that means we have to be really careful if we build something more intelligent than us to also have its goals aligned with our goals. For a silly example, if you have a very intelligent self-driving car with speech recognition and you tell it, “Take me to the airport as fast as possible,” you’re going to get to the airport chased by helicopters and covered in vomit. You’ll be like, “That’s not what I wanted,” and it’ll say, “That’s what you told me to do.”

SH: Right.

MT: But this illustrates how challenging it can be to get the goals right. There are a lot of beautiful myths from antiquity, going all the way back to King Midas, on exactly this theme, right? He thought it would be a great idea if everything he touched turned to gold—until he touched his dinner and then touched his daughter and got what he asked for.

And if you think about why we have done more damage to other species than any other species has on earth, it’s not because we’re evil. It’s because we’re so competent. Right? What about you, for example? Do you personally hate ants, would you say?

SH: No. And that’s a great analogy. It’s just that my disregard for them is often fatal to them. I’m so unaware of their interests that my mere presence is a threat to them.

MT: Right.

SH: As is our civilization’s presence to every other species. And here, again, it’s very hard to resist the slide into this not being just possible but inevitable.

MT: Right.

SH: The moment you admit that intelligence is just a matter of what some appropriate computational system does, and you admit that we’ll keep making progress building such systems indefinitely, unless we destroy ourselves in some other way—well then, at some point we’re going to realize, in silicon or some other material, systems that exceed us in every way and may ultimately have a level of experience and insight and form instrumental goals which are no more cognizant of our own than we are of those of ants. You know, if we learned that ants had invented us, that would still not put us in touch with their needs or concerns.

MT: That's right. And for an example above that, you actually know that in a certain sense, your genes have invented you. They built your brain so that you could make copies of your genes. That's why you like to eat—so you won't starve to death. And that's why we humans fall in love and do other things—to make copies of our genes, right? But even though we know that, we still choose to use birth control, which is exactly the opposite of what our genes want.

As you say, it would be the same with the ants. I think some people dismiss the idea that you can ever have things smarter than humans, simply for mystical reasons. Because they think that there's something more than quarks and electrons and information processing going on in us. But if you take the scientific approach, that you really are your quarks, then there's clearly no fundamental law of physics that says that we can never have anything more intelligent than a human.

We know that we were constrained very much by how many quarks you could fit into a skull and stuff like that, right? Constraints that computers don't have. It becomes instead more a question of time. And as you said, there's such relentless pressure to make smarter things, because it's profitable and interesting and useful, that I think the question isn't *if* it's going to happen, but *when.*

And finally, just to come back to those ants again, to drive home the point that it's really competence rather than malevolence that we should fear, if those ants were thinking about whether to invent you or not, one of them might say, "Well, I know that Sam. Actually, he saw me on the street once, and he went out of his way to not step on me. So I feel safe. I don't worry about creating Sam Harris."

But that would be a mistake, because sometimes you're jogging at night and you just don't see the ants. And the ants aren't sufficiently high up on your list of goals for you to pay the extra attention and see if there are any ants there before you put your feet down, right?

And suppose now you're in charge of this huge green-energy project. And just as you're about to let the water flood this hydroelectric dam that you've built, someone points out that there's an anthill right in the middle of it. Now, you actually know that the ants don't want to be drowned, right? And you have this decision. What are you going to do?

SH: Yes, well, too bad for the ants, in that case.

MT: Exactly. And we, I think, want to plan things ahead so that we don't end up in the role of the ants.

SH: Indeed. Well, listen, Max, now I'm excruciatingly mindful of the time, because you've been incredibly generous with yours. And there's a ton we could talk about. But I think this has been very useful, and I think our listeners will feel the same. To close, I want you to tell people where they can find more about you and your work online.

MT: People can go to my website or they can simply look for my book on Amazon, the one and only I've ever written. This book basically summarizes what I feel I've learned so far during my life as a scientist. And it's written for intelligent, curious people who have not spent their careers studying physics or other science.

I've tried very hard in writing the book to talk not only about the cool things you learn, but also about the joy of doing science, the process of it. That's why the book is subtitled "My Quest for the Ultimate Nature of Reality." So if you, the listener, choose to read this book, then it won't just be my quest, but our quest.

SH: And I highly recommend that they do. I'll put a link to the book and to your relevant websites on the blog post where I embed this podcast.

Listen, Max, thanks again. It's really been a pleasure talking to you.

MT: Thank you, Sam. It's been an honor and a pleasure. Thanks.

SH: Until next time.

[Closing Music]

If you enjoyed this podcast, there are several ways you can support it. You can leave reviews on iTunes or Stitcher or wherever you happen to listen to it. You can share it on social media with your friends. You can discuss it on your own blog or podcast, or you can support it directly, and there are two ways you can do this. You can support the Waking Up podcast through my website at samharris.org/support, or you can try a membership at Audible, the world's leading source of audiobooks, at audibletrial.com/samharris.

Notes

Find this article online at: https://www.samharris.org/blog/item/the-multiverse-you-you-you-you