NICHOLAS DIBELLA: So, last class, I began to introduce the mathematical universe hypothesis, which was proposed by a cosmologist here at MIT named Max Tegmark. And briefly, the hypothesis goes, while the universe is mathematical, more precisely, it's a mathematical structure. So, just to briefly, just to briefly review the arguments, this is the mathematical universe hypothesis.

Just to briefly go through the arguments, pretty simple. You start out by making the external reality hypothesis, which is simply that there exists an external world in addition to me, in addition to you. There is not just me in my head, in my mind. There's a whole external world out there. So this is just, this is the symbol for there exists. There exists an external world, exactly how it sounds.

You might choose not to believe this. You might choose to believe that your are the only thing that exists. Your mind is the only thing that exists. And that's actually, that's actually a very conservative viewpoint. I mean, it's quite a leap to say that other things exist. The only thing we know for sure exists is our heads. The only thing I know for sure exists is me.

The last class, we saw that rejecting this external reality hypothesis-- [INAUDIBLE]-- rejecting this hypothesis actually leads to kind of a complicated story. And so, last class we saw some arguments for them. And we generally like the simplest theory. The simplest theory is the one most likely to be true. So we rejected the rejection of the external reality hypothesis. We rejected solipsism, it was called.

And we saw that this external reality hypothesis is a reasonable hypothesis to make. Any questions about the external reality hypothesis? OK. So, if there is an external reality, then this external world should be intelligible to things other than us humans. If there exists intelligent extraterrestrials, then they would be able to learn about it. If it's possible to create artificial intelligence, then AI would be able to learn about this external world.

So this, so this external world is something that should be comprehensible to things that aren't necessarily humans, and therefore should be describable in a language that isn't distinctly human. I mean, it shouldn't depend on any concepts that humans have invented concepts, concepts like round or energy, and so forth. There might be some fundamental things, but it shouldn't depend on any baggage, we call it.

Baggage is just the distinctly human or distinctly alien language. So, the external reality hypothesis implies that external reality should be free from baggage. The description should be free from baggage. Now, something that is completely, something that can be, something that is a description of-- now, something that can be described in a completely baggage-free language, it's precisely what's called the mathematical structure. That's just what it is. And such a thing, such a thing is called a mathematical structure, i.e., the external reality is a mathematical structure.

Now, the more precise definition of mathematical structure is a set of abstract objects that are related in some kind of way. So, a set of abstract objects with relations. So, the external reality hypothesis implies that the external reality is a mathematical thing. It's a mathematical structure. So the universe is mathematical. So that's what the mathematical universe hypothesis is.

And so last class, I gave a simple example, a very simple example of a mathematical structure. I drew this thing, square. But what is a square, really? So, I've drawn something here, which is supposed to represent the idea, the abstract idea of a square. So what a square really is is it's a set of four things, we call them line segments. That's just what we call them.

And these line segments are related to each other in a certain way. One kind of way that they're related is that, well, each line segment is only adjacent to two other line segments. There's one line segment that isn't adjacent to it. Another property is that the way that they are adjacent to each other is by being adjacent in right angles. And the way that we represent this abstract idea is by drawing this. So this is a geometrical representation of the abstract mathematical structure that we know as square.

So that's a simple example of a mathematical structure. That's much more abstract. A square is much more abstract than we naively think of it as being. So, there are a lot of other mathematical structures. I mean, this is a very simple one. There are a lot more complex ones. There are also simpler ones.

Last class, I talked about the fact that all of our physical theories-- relativity, quantum theory, string theory, and so forth-- all these theories, they're, deep down, they're mathematical. They're mathematical things. And strictly speaking, what we do, strictly speaking, they're just a bunch of equations. I mean, relativity, for example, it's just a bunch of equations. That's all it is.

But what we do is we interpret those equations and we say, aha, they're actually, they actually

represent physical reality. Now, more precisely, the way that we interpret them is by saying, oh, well, we model the physical world as a mathematical structure. And the mathematical structure that we use is this. And that mathematical structure might be the mathematical structure of relativity or quantum mechanics or string theory or so forth.

And so the view that physicists generally take of the relationship between math and physics, the relationship between math and the physical world, is that the physical world can be described by math. The physical world can be described by a mathematical structure. But the mathematical universe hypothesis actually says much more. The mathematical universe hypothesis actually says that the universe is a mathematical structure. Not only is it described by it, it actually is a mathematical structure.

Universe is mathematical. Now, we don't know what that mathematical structure is today. Physicists are trying to figure out what it is. And we're trying to figure out what the theory of everything is. We're trying to figure out the theory of quantum gravity. But that theory corresponds to some mathematical structure. So we don't know what that theory is but people are surely working on it. Maybe we just haven't been clever enough or smart enough to figure out what it is.

But the mathematical universe hypothesis actually gives us hope in truly understanding reality. I mean, it could certainly be the case that reality is something that isn't mathematical. I mean, it could just be some weird kind of a thing. Who knows? Who knows what it would be? So, it could be the case that the true reality is just something that's not mathematical and it's just something that we can't understand.

Well, if it is mathematical, then, well, math is something that we know how to do. We've been doing it for a long time. And so we can just keep working on the math. You just keep trying to work to see what the right mathematical description of the world is. And then maybe one day we'll find it. One day, hopefully, we'll be smart enough or lucky enough to find it. And once we have that structure, then we'll really understand the true nature of reality.

Now, you may say, Nick, I'm looking at the universe. I'm just looking at the universe. It doesn't look mathematical. I mean, I mean, what's mathematical about a chalkboard or feelings? I mean, it doesn't seem like math. It doesn't seem like an abstract sort of thing. Right? I mean, that's the initial objection. That's an initial reaction that we would take.

Well, so it's actually, so to kind of answer this reaction, it's important to distinguish between two types of ways of looking at the worlds. So there's a bird perspective. There's a bird perspective of viewing the world, which is kind of like looking at-- you look at the mathematical structure outside of the structure, like a mathematician studying it. It's as though you're looking at the universe from above, like a bird.

So, there's a bird perspective. And there's also a frog perspective. These are Tegmark's terms. There's a frog perspective. In the frog perspective, you're an observer inside of the universe. And so, you're inside of it like a frog, being looked at by the bird. And so, an important question is how are these two, how are these two views related to each other?

So in the birth view, you have a mathematical structure and that's the universe. And so, in this structure, so you just have a mathematical structure. That's it. It's just math. But in this structure, there will be substructures. There'll be things like, things like trees, things like doors, chalkboards, and even humans. And these substructures have the property that they change in time and so forth. And there's a precise mathematical meaning to all these, to all these different properties.

Now, the special thing about the substructure that we are, the special thing about the human substructure, is that it possesses the property of being a self-aware, conscious substructure that actually perceives, that actually perceives other substructures. So it's because of this self-awareness property that we perceive the world as being physically real. Now, other less complex substructures, like a door or table, they just don't have perceptions. They're just too simple. They're not going to have any perceptions.

So you could imagine our universe being a less complex structure than it is. You can imagine it being simple enough that there are just no self-aware substructures. There are no conscious things inside of it, no conscious subsystems inside of it. In that case, the universe would be mathematically real. I mean, that's the only kind of existence there is, according to the mathematical universe hypothesis. The universe would be mathematically real, but there would be nothing there to call it physically real. There would be no appearance of physical reality.

So, in the mathematical universe hypothesis, mathematical existence is fundamental. Only when you have a complex enough universe, only when you have a complex enough mathematical structure that is the universe will you get any frog perspective at all. So

complexity leads to a frog perspective. In other words, complexity leads to the physical, to the appearance, the subjective appearance of physical reality.

So physical existence isn't a new, isn't a different kind of existence, according to the mathematical universe hypothesis. It's really a kind of existence that emerges from math. Now, last class I talked about, I talk about the relation between math and matter, physical existence and mind, mental existence. And I briefly looked at the idea that math is actually a product of the human mind. And the human mind is a physical thing. And so therefore, physical existence would give rise to mathematical existence.

Well, this is the opposite. Mathematical existence gives rise to physical existence. And also, mathematical existence gives rise to mental existence. So, in the mathematical universe hypothesis, the way that these three kinds of existences are related to each other is, so math is fundamental, leads the physical, leads to physical and mental.

OK. Are there any, are there any questions about what the mathematical universe hypothesis says and what physical reality is, according to the mathematical universe hypothesis? Related questions. Now, I know it's very abstract and it's a lot to take in. But, I mean, we're looking at really deep questions here and we should expect that the kind of answers that we would get would be weird or odd or make us feel a funny sort of way.

OK. So now, so if you believe, if you believe that mathematics is, in some sense, out there, if it exists independently of our own existence, if you believe that math isn't just the product of human mind but it, in fact, exists out there, the whole mathematical world exists out there, then you actually arrive at a new level of parallel universes.

I talked about it briefly at the end of the class where I talked about parallel universes. I guess that was five weeks ago. And this is called the Level IV multiverse, which is just that all mathematical structures are real. They're all real. They're all different universes on their own.

Actually, there's a slight technicality. Because you run into some paradoxes if you have all mathematical structures. So, it actually turns out, all computable mathematical structures. But that's a detail. Don't worry about that. So we'll just leave it like this, all mathematical structures are real. So this is actually an extra assumption. This is actually an extension to the mathematical universe hypothesis.

I mean, our universe is one mathematical structure. But this says that there are other

universes that are other mathematical structures. So this extension is actually called the ultimate ensemble theory. And this was also proposed by Tegmark. So this Level IV multiverse is actually the highest level in the hierarchy of parallel universes. So, in a Level IV multiverse, you have different mathematical structures that correspond to different universes. And each mathematical structure corresponds to different physics, different laws of physics.

So you start out at the Level IV multiverse. So this gives you your Level IV multiverse. You just have all these different universes, all these different Level IV multiverses. I'm sorry, all these different Level IV universes. I'm just, these circles aren't really meant to mean anything. I'm just drawing them. I'm just trying to represent the set.

So maybe we're living in this one. This is the one that has our set of laws of physics. This is the Level IV multiverse. And we're living in one of the members. We're living in one of the abstract mathematical structures. Last class, I also talked-- not last class, but when I talked about parallel universes-- question?

AUDIENCE: [INAUDIBLE]

NICHOLAS DIBELLA: Oh. There we go, Level IV. So, I talked about other levels when I talked about parallel universes five weeks ago. The top level, the highest level is Level IV, different laws of physics. So we have our own laws of physics. And from there, we get to the Level III multiverse-- I did the same thing-- which are, in a sense, different universes inside a given Level IV universe.

So we have three universes. And each one of these universes is supposed to represent a different part of true quantum reality. And it follows from the many worlds interpretation of quantum mechanics. The technical term is that each of these different worlds are different parts, different branches of the wave function of the universe. But that's just a funny phrase to use. But you, don't worry about that.

So maybe we're living in this, this branch of the wave function. The Level II multiverse that follows, and this one's the set of all universes with different physical constants and different space and time dimensionality. So, some of them have four dimensions of space. Some of them have 15 dimensions of space. Some of them have three dimensions of time, and so forth.

So you have all these different level universes that-- well, I describe them as coming from inflation, the cosmological theory of inflation. Maybe we're living in this one. Who knows? I'm

just trying to represent them.

And then the lowest level, Level I multiverse, which was the least controversial one. This was the set of all universes that start out with different initial conditions, different starting distributions of matter and energy. And ours started out with a given distribution, and we're living in it. So you have all these different Level I universes, and maybe we're living in this one.

So, we have a given law. We have a certain physics. We have a certain living in a certain branch of the wave function, a certain part of quantum reality. We have certain constants and we start out with certain initial conditions.

AUDIENCE: [INAUDIBLE]

NICHOLAS DIBELLA: Say that again.

AUDIENCE: Sorry. Basically, like Level IV, is it basically every universe has different mathematical structures?

NICHOLAS DIBELLA: Different mathematical structures correspond to laws of physics. Yeah

AUDIENCE: [INAUDIBLE]

NICHOLAS DIBELLA: Yeah. You can have funky laws of physics. Maybe you have, maybe you have a universe where-- yeah, like where gravity repels instead of attracts.

AUDIENCE: Yeah.

NICHOLAS DIBELLA: I made a horrible joke several weeks ago. I said, gravity is not responsible for people falling in love. Well, perhaps in another universe, gravity is responsible for people falling in love. That was a louder laugh than I got last time. OK. Yeah. So, yeah. Any, pretty much any conceivable physics that you can think of would be another world, just as real as our world, living in the Level IV multiverse.

And you can think of all crazy kinds of universes. Maybe, I talked about four fundamental forces, maybe there's seven fundamental forces. Maybe you like number seven, lucky seven. Maybe there is, so in our universe, we have the concept of space and time and forces and matter. Maybe there's a new kind of concept like, I don't know what you want to call it, ghosts.

Maybe there are ghosts or something.

If you can write down a consistent mathematical structure that has these properties that you want, then that structure would correspond to a real universe out there, living abstractly in the Level IV multiverse. We're living in an abstract universe according to these ideas. And there exist others in addition to us.

But only in those, only in those universes where, only in those universes that are complex enough to contain observers like us will there be the subjective perception of a physical reality. Otherwise, they're just devoid of observers and they're just math. They're just abstract things and there's nothing there to observe them. There's nothing there to call them physically real.

So, the mathematical universe hypothesis explains, as I talked about last time, it explains very well why math is so good at the describing the world. The world is mathematical, therefore it should be no surprise. Very simple explanation. Well, but you might still be wondering why are these laws of physics the laws of physics that rule our world.

In other words, why is this mathematical structure the one that is our universe, and not some other mathematical structure? Might still be possible. The mathematical universe hypothesis doesn't answer that. But the ultimate ensemble theory actually does, in a way. So, according to the ultimate ensemble theory, the Level IV multiverse exists. And there are many, there many other universes, as I just said.

So, the first thing to notice is that we shouldn't be surprised to find ourselves living in a universe where we exist. Because we exist, obviously, so it's something that shouldn't be surprising. Now, conceivably, the universe could be tweaked in a number of ways and we'd still be able to live in such a universe. So, conceivably, there are a lot of universes out there that contain observers that are just like us, that contain observers that are carbon-based, for example, that act like us, that have emotions, have intelligence, and so forth.

So conceivably, there could be a lot of universes out there in the Level IV multiverse that conceivably be our universe. So, the answer to the question, why this structure, not others, well, there's not really a good answer. Essentially, we live in this one. It was by chance. I mean, there are a lot of them that, there are a lot of them that contain observers like us.

So the ultimate ensemble theory actually makes a prediction. It predicts that we should find ourselves living in the mathematical structure that not only is consistent with our existence-- I

mean, it has to be consistent with our existence, otherwise we wouldn't exist in it-- not only that, but it should be, in a sense, typical among universes that we could exist inside of.

Now, what do I mean by typical? Well, so you look at the set of universes that we could possibly live inside of. Then what you want to do is you want to give each universe some probability of being the one that we live inside of. And if we live in a typical universe, then the probability should be something, I mean, you might have some, might have some probability distribution-- you look at universes this way, in this way is probability-- so typical might be something in this range. So you have a bunch of universes the fall into the typical range.

The ultimate ensemble theory predicts that our universe should be such a universe. It should be, in some sense, typical. So one day when we figure out, if we figure out the mathematical structure of our universe is, then we can simply say, well, let's look at this probability distribution and see if it's typical. If it is, then that's good evidence for the ultimate ensemble theory. If the universe turns out to be, if the mathematical structure turns out to be atypical or unusual in some sort of way-- if it follows, for example, in the tail of this probability distribution-- then that would be bad news for the ultimate ensemble theory because it predicts that we should be in a typical universe.

Now, a difficult question is how you determine this probability distribution, how you determine the probability of living in a given universe. And so that's probably the biggest problem facing this theory. Nobody really has any idea. Nobody really has any idea what the correct way of defining probability over these universes is.

But people have made suggestions. One suggestion that people have made is that maybe complexity is what defines this probability distribution. Maybe simpler structures are more likely to be the ones that we exist inside of, as opposed to more complex structures. It's similar to Occam's razor, where the simpler the theory, the more likely to be true.

So the simpler the mathematical structure, the more likely that we find ourselves living inside of it. So maybe these are the simple ones, the ones in the center of this distribution are the simpler ones, and these are the more complex ones. But nobody really has any idea if that's the right standard to use. And so this is an open question, how you define this probability distribution over universes in the Level IV multiverse.

It's actually also a big question in the Level II multiverse, how you define probability distributions for different physical constants and dimensionalities. So this is called the measure

problem, defining a measure over the universes. And it's currently unsolved for Level 2, and it's really, really unsolved for Level IV.

But if one day we have a compelling reason to choose one measure, one standard for performing this probability distribution over another, and then we calculate it and we find out one day what our mathematical structure it is, and we see that we're in a the tail end of it, we see that we're in a typical universe, then that would be bad news for the ultimate ensemble theory and we would have to go elsewhere.

But it's important that this theory is actually falsifiable. I mean, you can make predictions with it and you can test it with things you learn about the universe. That's always important for any theory that you want to take seriously. I mean, if the theory made predictions that you couldn't test in any kind of way, then the theory would be essentially useless. I mean, what would it be, what would it be good for?

OK. So that's the mathematical universe hypothesis and the ultimate ensemble theory. Mathematical universe hypothesis says that the universe is a mathematical structure. And the ultimate ensemble theory says that the set of all mathematical structures forms our complete reality. The complete reality is the world of mathematics. And the mathematical universe explains, this hypothesis explains why math is so effective, and this one explains why these particular laws of physics rather than others. Are there any questions about these ideas?

So, I just talked about one theory about the nature of reality, that reality is mathematical. More precisely, that reality is a mathematical structure. So now I'd like to talk about another one, which is that our reality is a computer simulation. And this is called the simulation argument.

AUDIENCE: Who was that proposed by?

NICHOLAS DIBELLA: An Oxford philosopher named Nick Bostrom. Simulation argument.

AUDIENCE: [INAUDIBLE]

NICHOLAS DIBELLA: Uh, well, actually the idea that our world is, in some sense, simulated or an illusion, goes back, goes way back. I mean, even Descartes, in his *Meditations,* he was using that reality is just a dream. Is it possible to know if you're living in, is it possible to know if you're dreaming or not? Is it possible, in principle, to know? Maybe the dream is just so vivid, could you really know if

you're living in a dream?

I mean, you might say, well, we're probably not living in a dream. But could you really know for absolute certainty? Maybe not. I don't think so. So the idea has been around for a while. But this argument, the simulation argument, it's pretty recent and really puts things in a modern perspective.

So the first thing, so the first thing to notice, the first thing to notice is that there are many people in the world, there are many researchers in the field of artificial intelligence. Why are you laughing?

AUDIENCE: [INAUDIBLE]

NICHOLAS DIBELLA: OK. OK. Is it my quotes? Never mind. So the first thing to notice is that there are researchers in the world working on artificial intelligence, the goal being to artificially create intelligence. So to build from non-biological stuff a program, computer program or a robot, some kind of machine that is in some way intelligence. For example, if you create an intelligent robot, you will be able to have an intelligent conversation with it. Maybe you talk about physics or world events or who knows. You could talk about a wide range of things.

It's also conceivable that you can, eventually one day-- at least these researchers believe-- one day you could eventually program artificial emotions. It's conceivable that sometime in the future we'll have robots for that, for example, they feel sad when they're neglected or they feel happy and fulfilled once they solve a hard, challenging math problem. So there's the prospect of artificial emotions.

Conceivably, you could also create a robot that had an artificial sense of humor, that would tell funny jokes that would make you laugh or maybe bad jokes that don't make you laugh but make you laugh afterwards because they're bad jokes. And so there's even, there's even the possibility that one day we'll be able to create artificial consciousness. And not only, not only a robot who's intelligent-- guys? Could do just turn it down a little?

So not only robots that are intelligent, not only robots that have emotions, sense of humors and so forth, but also are conscious, robots that are self-aware, that realize that they have a unique identity in the world that's different from the identity of all the stuff around them. So these are some of the ultimate goals of researchers in AI. We call it artificial because, because it's a biological thing. It's a machine thing. It's not carbon-based. It doesn't have anything really

to do with organisms. So the hope is, the hope is that, eventually, these things will be possible.

Most people in the field, most AI researchers, most cognitive scientists, and most philosophers of the mind believe that this program, this goal is one day reachable, that one day, we'll figure out how to do this, that consciousness is something that can emerge from non-biological stuff, that a mind can emerge from non-biological biological stuff. Once you just put it together in the right way, once you have the right architecture, the mind is something that can emerge.

So for example, just a few years ago, people built a computer, an intelligent computer that actually beat the world champion in chess. And chess is something that we generally associate with intelligence, right? People who are very intelligent are good at chess. That's something that we generally associate with chess. And when a computer beat the world champion in chess, that was a big triumph for AI research.

Now, you might say, well, that computer that beat the champion wasn't really intelligent. I mean, it was just doing all of these crazy calculations and figuring out all these possibilities and stuff. So it raises an interesting question that was actually raised when I talked about aliens several weeks ago. It's a question of what is intelligence. What defines intelligence? Does anyone have an answer what intelligence is?

AUDIENCE: [INAUDIBLE]

NICHOLAS DIBELLA: Right. So we have our own version of intelligence. So how might you define our intelligence, for example?

AUDIENCE: If you're trying to elevate [INAUDIBLE]

NICHOLAS DIBELLA: Yeah. So let's suppose our intelligence. So let's forget about the aliens for now. I mean, how would you define, how would you, how would you define our intelligence?

AUDIENCE: Probably to adapt and think, but also to actually pose questions. [INAUDIBLE]

NICHOLAS DIBELLA: So you define, you say something's intelligent if it can do these things, if it can ask questions, if it can adapt, if it can figure things out. If it can do these things, then it's intelligent. That's one prerequisite. That's one necessary requirement. OK. Any other? Yeah.

AUDIENCE: The ability to solve problems?

NICHOLAS Yeah. The ability to solve problems. So that's another thing that intelligent things should be

DIBELLA: able to do. That's the important verb. It should be able to do that thing. Yeah?

AUDIENCE: [INAUDIBLE]

NICHOLAS DIBELLA: The ability to learn. That's another thing that intelligent things should be able to do. Another one?

AUDIENCE: Memory.

NICHOLAS DIBELLA: It should have memory. It should remember things. Again, that's another thing that intelligent machines should be able to do. So, I keep emphasizing this word do because it emphasizes that an intuitive way of defining intelligence is simply in terms of behavior of something. So if a person acts intelligently, then we say he's intelligent.

And so, and so there's an idea proposed about 60 years ago by an English mathematician named Turing, called the Turing test. And this is a test to distinguish between intelligent and non-intelligent. So if you want to figure out if your machine is intelligent, for example, then the way that this task goes is you have a judge, a human judge. And you have a human judge in one room and you have a machine and a human in another room. And they're both communicating with the judge. They're having a conversation with the judge, for example.

If the judge, if the human judge is unable to tell the difference between the two, if he can't tell which is which, then-- I guess this is weird diagram, whatever-- if the judge can't tell the difference between the machine and the human, then the test says that, for all practical purposes, the machine is intelligent. If I talk to it and it acts like a human, it tells jokes, it makes insightful remarks, it figures things out, if it adapts to my personality, then we say it's intelligent. And so, if you can't tell the difference between the machine and the human, then the machine is intelligent. That's what the Turing test is.

So, it's important to have them in another room because, I mean, obviously, if you're looking at the robot, that's a good indication that it's a machine. I mean, you're looking at a thing that's doing something like that, it's a machine. I mean, unless it's a person trying to do the robot dance or whatever. Also, we talk to it. You probably also don't want to talk to it in person. You probably want to just communicate by text. I mean, if you talk to it and it sounds something like, a distant future, the year 2000, if it had a distinctive robot voice then that would be a good indication that the machine--

Maybe we don't yet know how to make the human voice. We can't have the distinctive human

voice. Maybe that was a bad human voice. I'm a bad human. Well, in any case, in any case, you want to filter out all these, filter out all these silly things and have kind of a text conversation, like an Instant Message, have them Instant Message each other, for example. And if the judge can't tell the difference between the machine and the human, then the machine is intelligent.

I don't think Smarter Child passes this test, unfortunately. But that would be a kind of Turing test. Smarter Child wouldn't pass the Turing test. OK. So this test emphasizes, I mean, this is kind of an aside. It's not really totally related to the simulation argument. It's an interesting aside. So the Turing test defines intelligence in terms of behavior. If it acts intelligently, then it's intelligent.

But you might object to that. You might say, well, I mean, you can have some, I mean, maybe it's just doing all this stuff and it doesn't really understand what it's doing. I mean, maybe it doesn't have concepts that we humans ordinarily use in figuring stuff out. Maybe it just tries all possibilities blindly, for example. I mean, we humans don't really think of things like that, like brute force, brute force types of solutions as being intelligent.

I mean, to do the brute force solution is intelligent, but actually doing it, you just blindly go through all these possibilities and try all this stuff out. We don't usually think it's an intelligent thing. So you might object to this. And you might also say, well, it doesn't really, I mean, it doesn't really test to see if the thing is conscious. It doesn't really test to see if the machine is conscious. I mean, the machine might be able to do all these intelligent things, but it might not have a sense of self. It might not have a sense of I.

And so defining consciousness is kind of a tricky business. But I think it's an interesting question, though, how you define intelligence. But nonetheless, most AI researchers and cognitive scientists believe that, in principle, it's possible to create a machine artificial intelligence and consciousness. You don't need a carbon-based, organism to have it. You could have a silicon-based system, like a computer. A computer is silicon-based.

So the simulation argument assumes what's called a substrate independence of consciousness. That's the first [INAUDIBLE]. A substrate is just a material. It doesn't depend on the material. So that's the first thing to notice. We'll assume that. We'll assume the substrate independence-- I'll just be a little more precise-- substrate independence of life. So we'll assume that. I mean, there's a minority of people, minority of philosophers of mind who

don't believe in this. But we'll take it, we'll take it for granted.

The next thing to notice is that computer power is increasing. Computers are increasing in speed and memory. And it's been like this ever since computers were invented. Computers have just continued to increase in computational power. Now our brains are very complicated things. And our brains are doing operations, many operations a second, dealing with very vast amounts of information, very complex kinds of information. So our brains are really, in a sense, computers doing all these things simultaneously, at the same time.

Now, our brain is actually a more powerful computer than anything that we have built for ourselves today, in a sense. I mean, even if we knew, even if we knew how the human brain worked, even if we knew how the mind and consciousness, all that stuff worked, we wouldn't be able to simulate the human brain today just because we don't have the computational power to. It's just so complex. In fact, it's been said that the brain is the most complex thing we know about in the entire observed universe.

But computational power is increasing. Computing power is increasing. So one day, in principle, we should be able to simulate the human brain, but once we figure out, once we figure out how to simulate artificial intelligence. I mean, I suppose if we simulate, if we want to simulate human intelligence, then that would be like artificial human intelligence. So that would be artificial intelligence, nonetheless.

But once we figure out how that all works, once we figure out how to build the right architecture so that consciousness would emerge, and then once we had the computational power to build a form of artificial intelligence, then it be possible to simulate the human mind. It would be possible to recreate the human mind. So the second assumption we make is that, one day, it's possible to simulate the mind.

And this really stems from the fact that AI researchers believe that one day they'll figure out AI, and also from the observation that one, well, that a computer power is increasing and that one day we'll have sufficient computer power to simulate the human mind. So that's the second observation we make. Now, why stop there? Why stop at one mind? Why not simulate another, simulate two minds or three minds or four minds? Why not simulate the entire human civilization?

I mean, conceivably, if you go far enough in the future-- maybe not 50 years, maybe not 100 years-- but maybe a million years, can you imagine that? Can you imagine what technology is

going to be like in a million years from today? It's just mind boggling. Who knows what technology is going be like in a million years?

So in a million years, this will probably be easy, to simulate a mind. And we'll want to do more challenging things, like simulating a whole society or the whole human civilization. So in fact, one day it may be possible to simulate the entire history of human civilization. I mean, and not only the history of humans, but also geological history and astronomical history and so forth. So one day it may be possible to simulate our entire reality if these two assumptions hold up and then we make the additional, make some further generalizations to this, simulate the mind, et cetera.

So one day, it's possible that we'll be able to create a computer program that contains self-aware entities who perceive themselves in a physically real world, you know, with all their surroundings that are familiar to us. There will be trees, grass. There'll be other people, other conscious things that are part of the program. So this is kind of like, if any of you have seen *The Matrix, The Matrix* starts out in the world that is, in fact, an illusion. I mean, all, I mean the main character is actually living in a computer simulation.

Now the premise is behind that simulation, like the super advanced real people, more real people need the humans to create power. That's a little silly. But nonetheless, that's the idea, that one day we'll be able to artificially create simulations of humans and human civilizations and all the human surroundings, trees, earthquakes, and the internet. And we can start at any point in the past. Maybe we'll decide to start at the point where man first invented the wheel, where man is extremely primitive. And then we can just let the simulation run.

We just let the program run. And then we could see how humans evolve, see how they progress. So humans will eventually grow as a population. Human knowledge will grow. And eventually, human technology will grow. And so, one day these humans in our civilization, if we're ever able to do this, if we're ever able to-- if these two postulates are true, sorry, these two assumptions are true-- that one day our simulations, the simulations that we produce may, in fact, want to run simulations of themselves. They'll be advanced enough to run simulations of themselves.

So then you have a simulated simulated reality. Then it can be the case that these simulated simulated humans are eventually progressed to the point where they want to run simulations, where they're able to, they're able to run simulations. So you can imagine having a hierarchy

of different levels of reality. So you start out at the true reality, the true physical reality. And then from there, the first civilization, the first civilization that emerges in the real world will run a simulation. So you have the simulated reality.

And then, eventually, they'll run simulations. And so you have another simulated reality. And then you can have another. And who knows how many you'll have? This could go on indefinitely, maybe infinitely. So you have this hierarchy of reality, starting out with the true physical reality, the top level, the non-simulated reality. Then you have the first simulated reality. Then you have simulations of simulations, then you have simulations of simulations of simulations, and so forth.

So you can kind of organize these according to the trueness of reality as it increases upwards. The top level would be the most real, the truest reality. The second level, the first simulated reality would be not quite there, but almost there. And then you have the second, which is less real, has less trueness of reality, and so forth.

Now in principle, it could be possible, to run simulations that are completely indistinguishable from the true reality. I mean, a sufficiently advanced civilization would be able to do that. I mean, right now we can't fathom it. But who knows in a million years? It might be easy to do. And it wouldn't even have to be a perfect description of the true reality. It could just be accurate enough so that the observers in the reality don't notice it. And even early on, you could start out having it not be very accurate at all. But then once the civilization progresses, you make it more, you increase the resolution, for example.

So in principle, it could be possible that each of these realities are indistinguishable from each other. Each simulated reality could be indistinguishable from the true reality. So now, you can ask the question-- so first, let's make the assumption. Let's make these assumptions. Let's assume the substrate independence of mind. Let's assume that it's possible to create AI. Let's assume that, one day, it will be possible to simulate the mind. And let's also assume that one day, we'll want to.

You know, one day, we'll want to. If this is the case, then you can ask the question, well, what's the probability that we're living in a simulated reality? To the best of our knowledge, I mean, since all these realities would seem the same way, they would be completely indistinguishable from each other. To the best of our knowledge, then, we can be in any one of these levels. There's no observable difference between any one of them.

So we should give each level, each one of these levels, just label the levels, give each one of these levels an equal probability. They're all the same, observationally. So give them all equal weight. So then you can ask the question, you finally ask the big question, what's the probability that we're living in a computer simulation?

Now, simply based on the fact that there's so many simulated realities, possibly infinitely many of them, and only one true reality, if you then ask what's the probability, it's very, very close to 100%. Because there are just so many of these. These just vastly outnumber the true reality. Each has an equal weight. So the probability is very, very close to 1 that we're living in a computer simulation. So that's the simulation argument.

You have to assume these assumptions to get the conclusion that we're probably living in a computer simulation. So if these assumptions are true, then we are probably living in *The Matrix,* or matrix-like thing. But you have to make those assumptions otherwise this is just a crazy argument. OK. does anyone have any questions about the argument the simulation argument. Not about the premises, not about the assumptions, but about the way that I formed the conclusion that we're probably living in *The Matrix?*

AUDIENCE: What's the odds that we're the true reality.

NICHOLAS DIBELLA: Very, very small. It's just 1 out of however many simulated, however many levels there are here. Yeah. So it could be 1 over infinity. So it could be 0, 0 probability. Yeah. Any questions? OK. Now, you might reject the argument by rejecting the premises. I mentioned before that there are some people who don't believe in the first one. Some people believe that consciousness can only exist in human form, for whatever reasons they have.

You can also object to the second one. Maybe it just never will be possible to run these simulations. I mean, maybe we'll one day understand AI, but we just won't ever have the computational power to do so, to run these simulations. Could also be the case that, well, one day we might have the ability but we destroy ourselves before we do it. I mean, that with great power comes great destructive capabilities. So, maybe our civilization is just doomed and it's just not possible.

So we can actually look at risks involved, existential risks, risks that threaten our very existence. And if we see that, if you do an analysis to see that we're very, very probably going to blow ourselves up or something like that, then that would be good evidence against the second assumption. So you could reject the second assumption for whatever reason that you

have.

AUDIENCE: [INAUDIBLE]

NICHOLAS DIBELLA: Either. I mean, the people could build it. Or maybe, eventually--

AUDIENCE: [INAUDIBLE]

NICHOLAS DIBELLA: Right. Right. Well, I mean, we don't know all the laws of physics. But I think we know, we think we know all the necessary laws for consciousness to emerge. I mean, when you look at the body, like when you look at how the heart works, you don't need to understand quantum gravity to understand how the heart works. So you don't need to know quantum gravity to understand why an apple falls down. You don't need the true, deep down, truly fundamental laws of physics to understand it.

The laws that we have right now are probably sufficient to figure out how the mind works, how the brain, how the consciousness and intelligence and emotion and so forth emerge from the brain. Yeah?

AUDIENCE: Well, if we destroyed ourselves before we could simulate reality, then wouldn't it be impossible for it to be simulated?

NICHOLAS DIBELLA: Well, I mean, so if it's probable for civilizations to eventually destroy themselves in some way before they have the ability to run these simulations, then yeah, that would be evidence against the conclusion. Then it would be the case that we're probably not living in the simulation. We're, in fact, living in the true reality. But who knows how likely it is that we'll destroy ourselves? It's hard to get a number like that. I mean, it's really hard to mathematize, it's really hard to quantify human actions and human psychology.

But the risk could be quite high. I mean, we definitely have the resources to do it. And if the wrong people get their hands on the right toy, the right weapon, then that's it. I mean, we could blow the world up many, many times. We have the weapons to do that. So some people even, some people actually, I mean, the probability that we'll destroy ourselves is a number that hasn't really been assessed very quantitatively by many people. But there are some people that give it a number as high as 1/3 that we'll destroy ourselves in the next century or so.

There's a high risk. There's a high risk involved. But, I mean, how do you actually get that number?

AUDIENCE: [INAUDIBLE]

NICHOLAS DIBELLA: Right.

AUDIENCE: [INAUDIBLE]

NICHOLAS DIBELLA: Yeah. I mean, historically, we've always figured out a way out of the doomsday, somehow.

AUDIENCE: We'll be OK.

NICHOLAS DIBELLA: Maybe.

AUDIENCE: [INAUDIBLE]

NICHOLAS DIBELLA: Yeah. Who knows? Who knows? So yeah, you might reject that second assumption. You might also reject the third assumption. Oh, a question. Yeah.

AUDIENCE: [INAUDIBLE]

NICHOLAS DIBELLA: Oh, yeah. So if it turns out that the simulated civilizations don't destroy themselves, then you're saying that would be good evidence for us that we're probably not going to destroy ourselves. Right, right. But we don't even know if these simulated realities exist. Yeah. I mean, if they did exist, that would be a good way of getting ahold of what the probability of our destroying ourselves is. I mean, often, scientists run simulations to get probabilities of things. Like how probable is this to happen? So we run a simulation.

For example, we simulate the weather. How probable is it that this is going to happen. So I mean, it would be useful. It would be useful to simulate the human civilization. Maybe we'll run a million civilizations of ourselves and we see that 250,000 of the civilizations eventually destroy themselves. That would be evidence that there's about a 25% chance that we'll eventually destroy ourselves.

AUDIENCE: [INAUDIBLE]

NICHOLAS DIBELLA: I'm not sure exactly I understand. I mean, the simulations would be completely removed. I mean, they would just be like a computer program that wouldn't be able to affect the world of the simulators. I mean, yeah. I mean, the computer program, you know, would keep running even after the civilization destroys itself. I mean, you can even imagine having time go by in the simulation twice as fast as it goes by here, or 100 times as fast. I mean, we want to know what's going happen to our future so let's simulate it. We don't want to wait. We want to simulate it and figure out what's going to happen. Then we we can get ahold of what will happen.

And so, I think it would be very, very cool if we're ever able to simulate human interaction. Because all the time I wonder what would happen if I said that. What would happen if I said that? What would happen, you know, if I did that? All the time I wonder about this sort of thing. So yeah, simulation would definitely help us to learn about our own future and our own nature. Yeah?

AUDIENCE: If we destroyed the universe that simulated us, would we cease to exist.

NICHOLAS DIBELLA: If we destroy the universe that simulated us, would we cease to exist? I suppose, because we exist inside the universe. So if we destroyed it, boom, that's us. Where do we go? But I mean, the programmers will be smart enough to avoid this sort of thing.

Like maybe one way that that could happen is to have a computer that like really super overheats. And then it could eventually become a black hole by the actions of the simulated civilization. Maybe that's one way of the simulation like destroying the universe that it's a part of. But if you're a smart simulator, then you probably want to program the simulation so that that sort of thing just doesn't happen. Yeah.

AUDIENCE: I have a question. You said that since there are so many simulated realities, that if each one made a new one. So wouldn't it be more likely that it's not a straight line? It's more like a branch, like maybe one civilization makes two.

NICHOLAS DIBELLA: Right, right, right.

AUDIENCE: Because if we were in like one of the higher ones, that means we having made another simulation yet.

NICHOLAS DIBELLA: Good question. I was actually going to just talk about that soon. Right. So I drew a pretty simple picture here. I mean, I said that there's a true reality, a civilization of true reality. Than they run a simulation. Then there's one simulation that runs another simulation. That runs a simulation, and so forth. So you have this straight line. But things will probably be more complicated than this.

I mean, you'd probably start out, I mean, if you started with your true reality, then you can imagine it being more than one branch. Maybe there were three branches. Then each one of these branches has its own branches. And then it keeps going on. Now you have a very, very complicated tree.

AUDIENCE: [INAUDIBLE]

NICHOLAS DIBELLA: That increases the odds even more.

AUDIENCE: [INAUDIBLE]

NICHOLAS DIBELLA: Right. So in this case, you just count the points in the tree. 1, 2, 3, 4, 5, 6, 7. You count them all and then you compare with the one true level, the true civilization. Then it becomes even more likely. But if you extend this idea even further, then it doesn't even work anywhere.

Because conceivably, you know, if the true universe is infinite, for example, then there will exist other civilizations. There will be other roots in this tree. There will be many branches and other roots, as well. And this will run simulations and then that will run simulations and you have another tree. If the universe is infinite, then you can imagine there being infinite number of roots because there are so many of them.

Well, in fact, you have an infinite number of civilizations in the true reality in an infinite universe, as well as an infinite number of simulated realities. And so then if we want to compare the two, give each civilization equal weight, each one equal probability, if you want to compare the two, then the probability that you get isn't something like, you know, 999999 over-- like, if there are 100,000 total, and 99,199 simulated realities, then it would be something like that, or that over infinity, whatever. It would, in fact, be infinity over infinity. That's the symbol for infinity.

And so what's that? What's infinity over infinity? So, it's no longer well-defined. And so this is actually a problem with the argument itself. So even if the assumptions are true, if the true

universe, if the true non-simulated reality is an infinite universe, then the probability isn't infinite. It's, in fact, if infinity were infinity, it's this weird thing. How do you deal with that?

Well, it turns out there are actually ways of getting around that. There are actually mathematical tricks of getting a meaningful number and getting ways of the argument, in fact, working. But it's kind of sophisticated and not really necessary for the next four minutes of class. OK. So, that's the simulation argument. It rests on these assumptions.

If they're all true, then we're probably living in a simulation. But you can actually form a general, you can actually form a more general conclusion, which is that either it will never be possible to run these simulations for whatever reason, or we'll never want to, or something like that, or we're probably living in a computer simulation. So there's really like a three-fold conclusion that you can make from this argument.

OK. So that's the simulation argument. Is it testable? Oh, I had some stuff to say about this. I'm running out of time. Maybe it's testable. I mean, certainly if you saw a dialog box appear suddenly in front of you and say, hi, you're living in a computer simulation, click to find out more, that would evidence for it. First, you probably think you're crazy. But if it happened many times to many people, then that would be evidence for it.

Anyone who has ever programmed knows that programs often have glitches or bugs or things like that. So if we're living in a simulation, if we're living in a computer program, occasional glitches could occur. In fact, it's been proposed that the apparent incompatibility between quantum mechanics and general relativity could just be a glitch in the program.

And so the first time that a distinctly quantum gravitational thing would happen-- I didn't talk about phenomena that would happen in quantum gravity, but one example would be a black hole evaporating. So the first time that quantum gravity would have to work in the simulation, the program might just crash. So the universe might just crash. So that's a problem.

So today, I looked at two theories of reality. Reality is mathematical, reality is a simulation. But I suppose even in the simulation argument, eventually you get to the true reality. Then you ask what's the nature of that. And in that case, the mathematical universe hypothesis would help you with that. But there are other theories. Maybe the universe isn't a mathematical thing. I talked about the relation between math, matter, and mind-- math, matter, and mind. They're related to each other. They're related to each other in some way, nobody really knows how.

But often in physics, when we see things that are related to each other, like space and time, they related to each other in some way-- in fact, Einstein figured out that they're unified in a single object-- it's been proposed that maybe math, matter, and mind are, in fact, different aspects of some fundamental thing. Math is some manifestation of this mystical thing. Matter is another manifestation of it. Mind is another manifestation of it. Who knows? That's another idea.

It's also been proposed, similar to the parallel universe stuff that I talked about, is that all possible worlds exist. All logically possible worlds exist-- A world in which I, you know, walked out of the classroom, a world in which I don't, a world in which I throw this up and don't catch it-- all possible worlds exist.

This is another theory that's been proposed, called modal realism-- I'm just trying to present as many interesting theories as I can in the time that I have left-- called modal realism. And it was proposed by a philosopher called David Lewis. And it's similar to these parallel universes that I talked about. But in fact, Lewis arrives at it through purely philosophical methods. He doesn't make any physical or mathematical arguments. In fact, he doesn't even believe that these possible worlds could be mathematical things. He arrives at it from a purely philosophical argument, which is interesting.

OK. So, summarize the course. Over the past eight weeks, we've talked about some big questions. Was there a beginning of time? Will there be an end? And so forth. And today I talked about the nature of reality. Although we're definitely a long way from knowing for sure what the answer to these questions are, I hope that I've been able to convince to you all that we're making progress. We've come a long way as a civilization, particularly over the last 100 years.

Now, these questions, when you first encounter them, they seem hopelessly difficult. They seem possibly impossible to answer. But with the rise of modern science and even philosophy, we've finally been able to attack these questions. And I think we've really been able to start to make some meaningful progress. Now, I don't know if we'll ever know for sure what the correct answers to these questions are, or even if it's possible to know in principle.

But I do find it uplifting that we minuscule systems, minuscule subsystems in a vast, possibly infinite universe, are able to even begin to unlock some of the deepest secrets of reality. And with that, it was a pleasure to have you all as a class. And I hope you enjoyed it as much as I

did. If anyone has any questions about any of the big questions that I've talked about, I'm happy to stay after class. OK. You're free to leave, but you have a question? Sure.

AUDIENCE: [INAUDIBLE]

NICHOLAS DIBELLA: How can you prove it? Well, the way that you prove these things about the universe, they're all by indirect means. I mean, in the first two weeks, I talked about the fact that we observe the universe is expanding. That was something that was noticed in the 1920s, 1930s. The universe is expanding. Distances between points in the universe is increasing.

Eventually, actually in 1998, it was finally observed that the expansion of the universe is actually accelerating. So not only is it expanding, but it's expanding at an increasing rate. So it's never going to stop expanding and eventually start contracting. It's, in fact, going to expand forever. So, we have observational evidence that the universe will expand forever, indefinitely, through acceleration caused by dark energy. And nobody has any idea what dark energy is.

AUDIENCE: [INAUDIBLE]

NICHOLAS DIBELLA: It's not going to run out. There's no reason to think that it's going to run out in any way. But yeah, if there's no longer any dark energy, than the universe would just be subject to the gravity of normal matter and dark matter and radiation, like lights, electromagnetic waves. In that case, the universe would eventually contract, eventually. But there's no reason to think that dark matter, that dark matter will run out. In fact, it seems to magically keep reappearing. The density of dark matter has been a constant over all the history of the universe.

I mean, usually the density of things decreases when you expand the volume. Like if you have a balloon that has a fixed number of atoms, then expanding the balloon will cause the density of the atoms to decrease. So the density of ordinary matter decreases when the universe expands. That's not the case with dark energy. When the universe expanse, the density of dark energy just stays the same.

AUDIENCE: [INAUDIBLE]

NICHOLAS DIBELLA: So it's really, so some people speculate that it's actually vacuum energy. There's always vacuum. There's always space. So there's always going to be energy from the vacuum.