Lecture Notes 8: The Nature of Reality

1 The Mathematical Universe Hypothesis

At the end of *Lecture Notes 7*, I began to introduce the Mathematical Universe Hypothesis, proposed by Max Tegmark, a cosmologist here at MIT. Briefly, the hypothesis is that the Universe *is* mathematical, or, more precisely, a mathematical "structure." He arrives at this conclusion through a very simple argument.

First, he assumes the External Reality Hypothesis, which is simply that there exists an external world that we are a part of: I am not all that exists; there exists an external world "out there." And we saw that this was a very reasonable position to take: not only is it intuitive, but it's also a *simpler* hypothesis than the hypothesis that there isn't an external reality (*i.e.* solipsism). The External Reality Hypothesis then implies that the external world should be comprehensible not only to us humans, but also to any extraterrestrial aliens (if they exist) and any artificial intelligence (if that can exist). So the external world should be describable in a language that doesn't depend on any of our distinctively human ideas, because who knows what kind of concepts an alien might use? So the external reality should be describable in a language that's free of any "baggage."

Now, something that can be described in a baggage-free language is precisely what's known as a "mathematical structure." A mathematical structure is simply a set of abstract objects with relations between them. Therefore, assuming the External Reality Hypothesis quickly leads to the conclusion that the Universe is a mathematical structure.

An extremely simple example of a mathematical structure is a square. A square consists of abstract things which we call "line segments" and the line segments are all related to each other in a certain way. Each has the property of being "adjacent" to two other line segments, and the precise way they're adjacent to each other is by being at "right angles" to each other.

These notions of “line segment,” ”adjacent” and “right angle,” are all truly abstract, but we can give them a geometrical representation and “draw” the abstract mathematical structure known as the “square.” [add figure of square]

There are many other, more complicated mathematical structures. Earlier, I mentioned that all of our physical theories about the Universe — relativity, quantum mechanics, string theory, and so forth — are mathematical in nature. Strictly speaking, what these theories *really* are is a bunch of equations; however, we *interpret* these equations as being a description of the physical world. More precisely, the way that we interpret them is by saying “Aha! We model the physical world by a mathematical structure, and the mathematical structure we use is this.” And that mathematical structure might be the one corresponding to relativity, or quantum mechanics, or what have you. And so the view physicists generally take about the relationship between math and physical world is that math describes the physical world; more precisely, the world is described by a mathematical structure.

But the Mathematical Universe Hypothesis says much more than this. According to the Mathematical Universe Hypothesis, not only is the Universe *describable* by mathematics — it actually *is* mathematical! As I've emphasized over and over again, we don't yet *know* what that mathematical structure is — whether it is related to string theory or loop quantum gravity, *etc.* — but the Mathematical Universe Hypothesis implies that we could truly understand our Universe by figuring out what that precise mathematical structure is.

It might be the case that we simply can't understand the true nature of reality. Maybe the Universe is something that isn't mathematical but rather of a nature that we humans couldn't even in principle understand. The Mathematical Universe Hypothesis implies that the Universe is a sort of thing we *can* understand; it's a mathematical thing. We humans know how to do math; we've been doing it for thousands of years. Maybe we're not clever enough or lucky enough today to have figured out the mathematical structure of the “theory of everything,” but the Mathematical Universe Hypothesis gives us hope that, one day, if we ever we figure out what that structure is, then we'll finally understand the nature of reality.

Now, you may say, “Nick, hold on. I'm looking at the Universe right now, and it doesn't *look* mathematical! How can you say that all of this is just some kind of an abstract *thing*? I feel real, I'm alive!” Well, let's try to be clear about things. We need to distinguish between two different ways of viewing the Universe: the *bird perspective* or "outside" view, of a mathematician studying the mathematical structure, and the *frog perspective* or the "inside" view, of an observer living inside of it.

In the bird view, when you look at the Universe, you simply see a mathematical structure. And, in this structure, there are substructures. One substructure may be what we call a "tree," another may be a "MacBook," and another might be a "human." The special thing about the human substructure is that it is complex enough to possess the property of being *self-aware*. It is because we humans are self-aware that we *perceive* the Universe as being "physically real." Our complexity gives rise to our "frog" perspective. Other, less complex substructures are just too simple to have "frog" perspectives.

It's easy to imagine our Universe being a less complex structure than it actually is. If the Universe were sufficiently simple, there could still be substructures, but the universe would simply be too simple to have any self-aware substructures. In that case, our Universe would merely be "mathematically real" as opposed to "mathematically real and physically real," which the Mathematical Universe Hypothesis implies our *actual* Universe is (because we humans exist in it and perceive it).

So, according to the Mathematical Universe Hypothesis, mathematical existence is fundamental. More precisely, according to the MUH, there is only one kind of existence, and we call that kind of existence "mathematical." *Physical* existence is only something that emerges in certain mathematical structures. However, there's nothing fundamentally *different* about physical existence, from the *bird* perspective. A structure is just a structure, to a mathematician. The notion of "physical existence," is merely a *label* the "bird" uses to distinguish among structures that have self-aware substructures and structures that don't. A structure is "physically real" if there is any frog perspective at all.

2 The Level IV Multiverse

Now, if you believe that mathematics exists, in a sense, "out there," then you actually arrive at another level of parallel universes, in addition to those that I talked about in *Lecture Notes 3*. This is the Level IV Multiverse: the set of all mathematical structures. Our Universe is a particular mathematical structure — we don't know yet what it is — and it's just one structure among the infinitely many mathematical structures. Other structures are literally other "universes." However, only in structures that are sufficiently complex to allow for the existence of self-aware substructures will there be the subjective appearance of a "physically real" world. This idea that all mathematical structures are, in a sense, "universes" is actually an extension to the Mathematical Universe Hypothesis, and is known as the "Ultimate Ensemble Theory" (also proposed by Tegmark).

The Level IV Multiverse forms the highest layer of the multiverse hierarchy. At the top level are "universes" with differing laws of physics; these are the Level IV universes. At the next level, Level III, are different "branches" of the wavefunction of the universe, *i.e.*, different "worlds" of the true quantum reality, according to the many-worlds interpretation of quantum mechanics. Next, at Level II, are universes of differing physical constants and dimensionalities — where, for example, the number of space and time dimensions vary, the speed of light varies, and so forth. Finally, at the lowest layer, are the Level I universes. These are the universes with varying starting conditions for the distribution for matter and energy.

In *Lecture Notes 7*, I raised the question of why math is so effective at describing reality. The Mathematical Universe Hypothesis gives a very simple answer to the question, the answer being that the Universe *is* mathematical! You may still be wondering, though, why our Universe is the particular mathematical structure it is. After all, there are a lot of structures out there, so why is our Universe this one? In other words, why are the laws of physics the way they are? Conceivably, things could have been different, so why this way?

In view of the Level IV multiverse, this question isn't actually that hard (in principle). First of all, we shouldn't be surprised to find ourselves living in a universe where we exist because, obviously, we exist! The answer, then, is simply that our universe is located somewhere in the Level IV Multiverse — it's a universe that's consistent with everything we know about our universe. There exist things like space and time, forces, matter, and so forth. So, there's not really a good "reason" why we live in this universe; we're basically in this one by chance. We could've been in any one that's consistent with our existence.

The Ultimate Ensemble Theory actually predicts that, of these universes that are consistent with our existence, we should find ourselves living in a universe that is in some sense "typical" of these universes. What "typical" means is actually a difficult question that many people disagree about — this is probably the biggest problem for the theory — but it's important that the

Ultimate Ensemble Theory is actually falsifiable in principle. It's falsifiable in the sense that, if you looked at all the universes that are consistent with our universe, and you find that the mathematical structure of our Universe is somehow *a*typical or *unusual*, then that would be a contradiction between prediction of the theory and observation of what our mathematical structure actually is, and would therefore bad news for it. But the theory explains *why* math is so effective and also *why* the laws of physics are what they are.

OK, so that's the Mathematical Universe Hypothesis and the Ultimate Ensemble Theory. The Mathematical Universe Hypothesis says that the Universe *is* mathematical. *That's* the nature of reality — math. The Ultimate Ensemble Theory says that the *complete* reality is the set of all mathematical structures, in other words, that the complete reality is the complete world of mathematics.

3 The Simulation Argument

Now I'd like to discuss another theory about the nature of reality, known as the Simulation Argument, which was proposed by Oxford philosopher Nick Bostrom. We start out by making some observations.

First, we notice that there are many people in the world working in the field of AI, artificial intelligence. These AI researchers are trying to create intelligence artificially, the ultimate goal being to create some sort of computer program or a robot that would be, in a sense, "intelligent." For example, if you created an intelligent robot, you'd be able to have an intelligent conversation with it — perhaps you'd talk about physics with it, or world events. It's also quite conceivable that you could create a robot that would some kind of artificial emotions — maybe it would become "sad" when it feels neglected, or extremely happy and fulfilled when it solves a hard math problem. Maybe it's possible to program a robot to have a sense of humor, who can tell funny jokes. And *maybe* it's even possible to create artificial *consciousness* — not only a robot who can do intelligent things and feel emotions, but a robot who is also *self-aware*, who realizes that it has a unique identity which is different from the identity of everything else in the world.

These are the ultimate goals of researchers in artificial intelligence. We call it "artificial" because it's not a biological kind of intelligence; it's a machine intelligence. However, the hope of AI researchers is that things like intelligence, emotions, and even consciousness can *emerge* once you've built stuff in the right way. Nobody in the world knows *how* exactly you're supposed to do this, but many researchers feel we are making progress. For example, just a few years ago, people built a computer that *beat* the world champion of chess. It's debatable whether this is evidence of artificial "intelligence," but what is intelligence, anyway?

One test was proposed about 60 years by the English mathematician Alan Turing. It goes as follows. You have a human judge, who engages in conversation with one human and one machine, each of which tries to appear human. If the judge is unable to tell which is which, then the machine is said to pass the test. Now, obviously, you'd want to have the judge in another room — otherwise, he'd easily be able to tell which is which — so a good way of doing the test would be, for example, to have both the human and the machine instant message the judge. If the judge can't tell which is which, then the machine passes the test and is said to be "intelligent." (I don't think SmarterChild passes this test, unfortunately.)

So, this test — the Turing test — defines intelligence purely based on behavior. Maybe that's not the best way to define intelligence. Also, even if the machine is deemed "intelligent," it may not be *conscious*. So, how to test for *consciousness* is trickier. Nevertheless, most AI researchers and cognitive scientists today believe that, in principle, it *is* possible to create non-biological intelligence and consciousness, i.e., they believe in the "substrate independence" of consciousness. You don't need a carbon-based organism for consciousness to emerge; if you're smart enough, you can implement consciousness on, e.g., a silicon-based system, like a computer, given the right computational architecture.

So, the simulation argument assumes the substrate independence of consciousness. There are some (a minority of) philosophers of mind who don't believe in this, but let us assume it is true.

The next thing to notice is that computer power is increasing. Computers are increasing in space and in speed. This is something that's been true ever since computers were invented. Now, our brain is a very complicated thing, perform vast and complex amounts of operations per second. With today's technology, even if we knew how to simulate the types of processes going on in the brain, we wouldn't be able to, because our technology isn't sufficiently advanced. However, if computing power continues to increase, then in principle, at *some* time in the future, we will be sufficiently advanced enough to implement a human mind on a computer. At some time in the future, it's possible that we'll be able to artificially create all of the kinds of processes that are going on in the human brain. In other words, it's possible that eventually we'll be able to *simulate* the human mind.

But why stop there? Why not simulate two human minds, or three human minds, or an entire *civilization*? This, too, may be possible in principle once we're sufficiently technologically advanced. In fact, it may be possible one day to simulate the entire history of human civilization — not only the history of humans but also the history of everything *else*, including geological history, astronomical history, and so forth. So, it may be possible one day to simulate our *entire* reality. We may, one day, be able to create a program that contains self-aware entities who perceive themselves in a "physically real" world as their surroundings — things like trees, earthquakes, and the Internet.

And we could start at any point in the past. Maybe we'll choose to start from the point where man is extremely primitive and has just invented the *wheel*, for example. Then we could run the simulation and let the civilization progress. Eventually, the human population would grow, human knowledge would grow, and so would human technology. Eventually, this simulated civilization will have the technology to simulate *themselves*. And, if they choose to do that, then this *simulated* simulated civilization will eventually have the technology and possibly desire to simulate *themselves*. And the process would continue indefinitely.

So, we can imagine starting at the civilization that was the *first* to run simulations of themselves. We can call their reality the "true" reality. Then they run simulations and the reality of their simulations is a "simulated" reality — a "less true" reality. And then the reality of *their* simulations is an even "less true" reality. And who knows where this would stop?

So, let's make the assumption that substrate independence is true, and also the assumption that simulation of human civilization is computationally possible. Furthermore, let's make the assumption that once such simulations are possible for a given civilization, the civilization (or possibly even a single individual in the civilization) has the desire to *run* simulations. *Then*, this hierarchy of simulated realities would emerge.

Now, remember, each of these simulated realities seems "real" to the entities part of the simulations; the simulations could be programmed that way. Now, you can ask the question: what's the probability that we're living in a simulation?

Well, if each simulated reality is indistinguishable from the "true" reality — the top level in this hierarchy — then, to the best of our knowledge, our reality could be any one of these levels. There's no observable difference among any of them. So we should give each level an equal probability. And, because there are simply more simulated levels than non-simulated levels — there could conceivably be *many* simulated levels, as opposed to the 1 non-simulated level — this would then mean that the probability that we're in one of these simulated levels is very, very high, very close to 100%.

So, if these assumptions are true, then this argument — the simulation argument — shows that we are *very probably* living in a computer simulation.

Now, of course, you may object to these premises. The first you may object to on the grounds that only humans, for example, can possess consciousness, for whatever reasons you have, as some people have done. You may also reject the second that we'll ever become sufficiently technologically advanced. You may have your own reasons for that; perhaps we are doomed to destroy ourselves by our technology. You may also reject that, once we have the ability to run these simulations, we may not have the desire to; maybe there are some laws in the future against this sort of thing. But this objection would have to hold true for *every* individual in every civilization, which I think is a rather hefty psychological assumption. In the future, it may be extremely simple for a single individual to run a simulation of a civilization. And there are bound to be curious people, people who may break the law, and so forth.

You may also have a problem with the logic of the argument itself. ["tree" diagram of levels of reality...], showing how one simulation may merge from another at a bunch of levels. But things are bound to be more complex than this. Because maybe, for example, the first civilization doesn't just run one simulation but in fact runs several, or maybe 100. So you can have a number of complicated branching scenarios. But all of these assume that you start out with a single root, a single civilization existing in the "true" reality. You could conceivably start out with more than one root, if, for example, there exists intelligent *aliens*. And, conceivably, if the true reality is an infinite universe, you could start out with *infinitely* many roots. If *that's* the case, then if you assign equal probability to each "reality" — each corresponding to a point on this tree — then, since you've got an infinite number of points for the "true," non-simulated reality, and an infinite number of points for the simulated reality, then you get an ill-defined answer for the probability of us living in a simulation. Essentially, you get an infinity divided by another infinity — what's that supposed to be? So, that's one way the argument itself may fail to work, although there are actually mathematical ways of getting around this.

3.1 Is the Simulation Argument Testable?

Can the simulation argument be tested? Could we ever know if we're living in a simulation?

What would you think if, al of a sudden, a big dialog box appeared in front of you, which had the message, “Hello, my child. You are living in a computer simulation,” before disappearing? Well, at first, you'd probably think you're crazy, but if this sort of thing legitimately happened many times to many people, then I suppose in principle that would be evidence that we're living in a computer simulation. Or perhaps we could be uplifted to our simulator's level of reality. It's kind of silly, but I think that would be evidence for it.

However, maybe our simulators don't want us to know that we're simulations, or, at least, they don't want to make it easy for us to know. Then could we know? Well, if we ever get to the point where we're about to run simulations of ourselves, then I think that would definitely be very strong evidence that these two assumptions are true, forcing you to accept that we're probably living in a simulation. Unfortunately, we're definitely *not* at that point technologically, so we can't use this method to “test” the Simulation Argument? Is there *any* way we could test it?

Well, maybe. Anyone who's ever programmed before knows that “glitches” can sometimes occur. Maybe there's a bug in the program, or something is wrong with the hardware. So, if we're living in a computer simulation, we might expect to sometimes see glitches in our world. In fact, it *could* be that the apparent incompatibility between general relativity and quantum mechanics is just a bug in the program! Why bother trying to come up with a theory of “quantum gravity”? It's doomed to fail! In fact, it could be that the first time a distinctly quantum-gravitational effect happens — I haven't talked about any such effects, but one effect would be the evaporation of a black hole — it could be that once, something like this is about to happen, our simulators program would crash! The Universe would crash! (Then, our simulators would know that they shouldn't have upgraded to Universe Vista . . .)

OK, so that's the Simulation Argument. If these assumptions are true, then it follows that we are very probably living in a computer simulation. It's a tricky business to test this argument, but it might be possible.

3.2 Other Theories

Although the Simulation Argument says that *our* reality may not be the "true" reality, you could still ask questions about the nature of this *true* reality — what is the *true* reality? And that's where something the Mathematical Universe Hypothesis would come in. There are, of course, other answers people have given besides the Mathematical Universe Hypothesis. Probably the most common answer is "who the heck knows?" But there are, of course, more serious answers that people have given.

For example, last class I talked about different "kinds" of existence that people talk about — physical existence, mathematical existence, and mental existence. And there seems to be interrelationships among them. Well, you can ask the question: well, what exactly *is* that precise relationship? The three kinds of existence seem related to each other, but how? I suppose the number of different theories you can come up with is virtually infinite, but a particularly interesting one is as follows. Perhaps math, matter, and mind are really unified and are aspects of an *underlying* aspect of reality; they're different manifestations of an as-of-yet unknown kind of reality.

This concept of unification is similar to something that has historically happened all the time in physics. Electricity and magnetism, for example, initially seemed like two very different things. However, it was eventually noticed that they were related to each other, and eventually it was discovered that the most elegant way of viewing electricity and magnetism are as different manifestations of a single *electromagnetism*. A similar thing happened with space and time in Einstein's theory of relativity. Perhaps the same thing will one day happen with math, matter, and mind.

And there are other theories out there. For those of us who think the idea of "possible worlds" is cool, there was a theory proposed by the philosopher David Lewis known as "modal realism," according to which *all* possible worlds exist — a world where I eventually get tired and fall asleep while teaching, a world where I leave the classroom, and so forth. Also all possible worlds of different physical constants, dimensionalities, and even laws of physics. And these worlds all have the same "kind" of existence — you can call it "physical" existence if you want. So, essentially, the Level IV Multiverse of Tegmark (although Lewis actually did not believe that the possible worlds could be mathematical in nature). The important difference between the two theories is that, while our modern-day ideas of "parallel universes" stem from physical theories and mathematical considerations, the theory of modal realism is based on purely *philosophical* arguments, which makes it even more interesting.

Over the past 8 weeks, I've talked about a number of "big questions." Was there a beginning of time? Will there be an end? Is the Universe infinite? Are there parallel universes? Do aliens exist? Is time travel possible? What's the theory of everything? Is a theory of everything even possible? And what is the nature of reality?

Although we are definitely still a long way from knowing for sure what the answers to these questions are, I hope I've been able to convince you that we've come a long way as a civilization, particularly over the past 100 years. These questions, when you first encounter them, seem hopelessly difficult. And, for many, many years they were pretty much impossible to answer. However, with the rise of modern science, we've finally been able to attack these questions, and I feel we've really been able to make 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 in principle to know. However, I do find it uplifting that we — miniscule systems — in a vast, possibly infinite Universe are even able to *begin* to unlock some of the deepest secrets of Reality.