Lecture Notes 7: Musings on the Theory of Everything and Reality

1 Is a "Theory of Everything" Even Possible?

In *Lecture Notes 6*, I described the ongoing quest of theoretical physicists today in uncovering the "theory of everything" that describes Nature. Nobody knows yet what that theory is — it would have to somehow reconcile the conflict between quantum theory and general relativity, *i.e.*, it would have to be a theory of "quantum gravity" — but people are certainly trying, and many optimistic people even feel that we're "almost there." I personally have no idea if we're close at all, because I am definitely *not* an expert on any candidate theory of quantum gravity. However, I do feel that we've come a long way in the past 100 years. And, when I marvel at the progress we've made, it makes me feel that, *some day*, we will finally have the correct theory of quantum gravity. Whether that day is 50 years from now or 500 years I won't try to guess. But it seems like we *can* understand the Universe. Indeed, as Einstein once mused: "The most incomprehensible thing about the Universe is that it is comprehensible."

But is the Universe *really* comprehensible? We can certainly understand aspects of the Universe today, but could we someday understand *everything* about it? The prospect of understanding everything — not only the laws of physics, but also the state of every single object in the (possibly infinite) Universe — probably seems rather bleak. After all, we humans are only only small, finite systems in an incredibly vast (possibly infinite) Universe, so maybe complete knowledge of every system in the Universe is impossible to attain. But perhaps we can still completely figure out the *rules* that describe these systems, *i.e.*, the laws of physics.

Maybe. But first, let's take a step back. One of my goals for this class has been to explain the best answers physicists have to big, foundational questions at a purely conceptual level. And I think I can convey the *sense* of these ideas purely conceptually. However, deep down, these physical theories that I've been describing — general relativity, quantum mechanics, string theory, and so forth — are mathematical in nature. Strictly speaking, these theories are nothing more than a bunch of mathematical equations that physicists *interpret* as explaining the world. And, because these theories can be pretty complicated mathematically, I've refrained from trying to explain them at a mathematical level; I haven't assumed any mathematical knowledge other than knowledge of what multiplication is.

But it's important to understand that all of our theories of physics are actually precise, mathematical theories. Everything I've described in all of these notes and in the lectures can be traced back to precise, meaningful mathematical notions. So, if anything I've said has sounded vague or unclear, you'll just have to take my word that everything actually makes good sense when you study the theories in their full mathematical detail.

1.1 Gödel's Incompleteness Theorem

I'd now like to describe one of the most surprising results discovered in modern mathematics, and I'll soon turn to how it applies to the theory of everything. But first, the question: what is mathematics? Most of us, when we've first exposed to math, don't really see it as much more than a set of tools or a bag of tricks. If you want to describe *one* situation in the world, you use a certain set of tools — algebra, for example — and if you want to describe *another* situation, you use some other tools — geometry, say.

But this isn't the way that a pure mathematician would view mathematics. To a mathematician, mathematics really has a life of its own, kind of an abstract form of existence. We may see certain mathematical objects in our everyday life — various shapes and numbers, for example — but mathematicians also study many objects that don't seem to appear anywhere in the real world.

A simple example is imaginary numbers, to be distinguished from "real" numbers. Real numbers are the numbers most familiar to us. They include the integers — whole numbers like 4, 7, and 0, and their negatives like -4 and -7 — as well as non-integers — numbers like 1.9, -8.642, and 14.2194. And a simple thing we learn early on is that the *square* of a real number — the real number multiplied by itself — is always a positive number (or zero), since a positive number times a positive number is a positive number, and a negative number times a negative number is also a positive number. Thus, every real number is the *square root* of some positive number (or zero). For example, 3 is the square root of 9, since $3 \times 3 = 9$, and 4 is the square root of 16, since $4 \times 4 = 16$.

What's the square root of -1? Well, it can't be any real number, because the square of every real number is greater than or equal to 0. However, there's nothing stopping me from simply *defining* a "number," which I designate by the symbol i, which has the property that $i \times i = -1$ (so that $2i \times 2i = -4$, $1.5i \times 1.5i = -2.25$, *etc*). Such numbers are called "imaginary" numbers to distinguish them from the real numbers.

Now, when we look at the world around us most of us don't see imaginary numbers anywhere. We certainly don't count in units of imaginary numbers, and we definitely don't measure things to have imaginary quantities. So the imaginary number system seems to be a purely abstract thing that doesn't find its place in the real world. And there are many other purely abstract entities that mathematicians deal with that don't seem to have anything to do with the real world. Imaginary numbers are just a particularly simple example to describe.

Now, all of mathematics is based on a certain number of "axioms" or "postulates." These are some fundamental premises from which one can derive a number of consequences of mathematical truth. For example, if you've taken a geometry class, you might have heard of Euclid's axioms, which describe shapes, lines, points, angles, *etc.*, that lie on flat surfaces (*i.e.*, plane geometry):

1. Any two points can be connected by a straight line.

2. Any straight line segment can be extended to become an infinitely long line.

3. Given a straight line segment, a circle can be drawn having the segment as radius and one endpoint as center.

4. All right angles are congruent.

5. *Parallel postulate.* Parallel lines never intersect.

From these fundamental axioms, you can prove more geometrical facts, or "theorems." For example, you can prove things like the Pythagorean theorem, the fact that the area of a circle is πr^2, and so forth. And, if you're smart enough, there's no telling how many brilliant theorems you can derive about triangles, polygons, circles, and so forth.

OK, so much for geometry. Similarly, you can write down some fundamental axioms for *arithmetic*, that is, the math having to do with whole numbers — the addition and multiplication of numbers like 0, 1, 2, and so forth. The axioms are a bit more complicated, so I won't list them for you, but the story is the same. You start off with some axioms from which you can derive a number of theorems; that's pretty straightforward. For example, you can show that an odd number plus an odd number must be an even number, that there are no even prime numbers greater than 2, *etc.*

Now, Dear Reader, prepare to have your mind blown. It turns out that there exist true statements about arithmetic that simply *can't be* derived from the axioms of arithmetic. There are some arithmetical truths that simply can't be proven. In a sense, then, *truth is more fundamental than provability*. This mind-boggling fact about arithmetic is known as *Gödel's incompleteness theorem*, after the German logician who proved it. *Why* this theorem is true is a story for another day, but for now I simply ask you to take my word for it (or Wikipedia's word, or the word of a million other websites).

Gödel's theorem is actually a bit more general than this. In fact, the theorem applies to any mathematical system that

1. has a finite number of axioms,
2. is "large" enough to include arithmetic, and
3. is consistent.[1]

Euclidean geometry, for example, has a finite number of axioms: 5. It is also consistent, since none of the axioms contradict each other. However, it turns out that Euclidean geometry isn't large enough to include arithmetic. So, Gödel's incompleteness theorem does not apply to Euclidean geometry, and everything about Euclidean geometry — points, lines, triangles, *etc.* — can be proven from the 5 axioms.

[1]A system is consistent simply if the axioms don't contain any contradictions. For example, if one added the extra axiom to Euclidean geometry that "not all right angles are congruent," then obviously this new axiom would contradict axiom 4. Euclidean geometry, appended with this new axiom, would therefore be inconsistent.

Now what about the theory of everything? We assume it's mathematical in nature — mathematics seems to be the language in which Nature speaks — so it should be built out of some kind of mathematical system, which in turn is built on some axioms. So, the question is: does Gödel's theorem apply to this mathematical system? If so, then there would exist true things about the Universe that simply *aren't* provable. In other words, there would be things which we could observe in the world but which we would be at loss to "explain" in the same way we explain other physical phenomena, by starting with fundamental principles and then working our way up.

For example, the way we explain why apples fall from trees is to give the following simple argument: The apple has the property of having a thing called "mass," and the Earth also has the property of having the thing known as mass. A fundamental law of Nature — the "law of gravity" — is that objects with "mass" *attract* each other. Therefore, once the tree is no longer holding onto the apple, the law of gravity kicks in, and the apple and the Earth attract each other, causing the apple to fall down onto the Earth. Thus, we have taken something we observed in the world — apples falling down from trees — and "explained" it by appealing to a fundamental "axiom" of Nature — the law of gravity.[2] If Gödel's theorem applies to the true theory of everything, then there would exist facts about the Universe that simply *can't* be explained in the way that we explain why apples fall from trees.

So, let's see what Gödel's theorem has to say about the theory of everything. First, will the mathematical system that describes the theory of everything have only a finite number of axioms? Well, every physical theory that has ever been written down has, so it seems natural for the theory of everything to as well. But maybe not; Gödel's theorem wouldn't apply, then, and the theory would be "complete." (Although, if the theory requires an infinite number of axioms, could we even figure it out? And, among a host of other issues, where would we store our *knowledge* of it?)

Second, will the axioms of the theory of everything turn out to be consis-

[2] Although he didn't claim to have a "theory of everything," the law of gravity would be an axiom in Newton's theory of gravity. Quite remarkably, we now know that this "axiom" isn't really an axiom at all, because it is derivable from even more fundamental principles, namely the axioms of Einstein's general theory of relativity. In general relativity, the fact that objects with mass attract each other actually *emerges* in an elegant way from the intrinsic curvature of spacetime. However, *why* there exists a thing called "spacetime" that possesses the property of "curvature" is left as an unexplained (perhaps unexplainable) axiom of general relativity.

tent? I certainly hope so. How can a fact about the Universe be simultaneously true and not true? I personally have a very hard time understanding how to extract meaning from a contradictory statement, and I've never met anyone who has claimed to "understand" contradictions.[3] But maybe the axioms will, in fact, turn out to be inconsistent (even though there's absolutely no reason in the world to suspect that they will). So, an inconsistent theory would also result in a "complete" theory.

Conditions (1) and (3), then, seem like they should probably hold for the theory of everything. The most uncertain condition would seem to be condition (2).

Unfortunately, as I've emphasized, we don't yet know what the correct mathematical system to describe the theory of everything is. It could be that the system is not large enough to include arithmetic. Whole numbers seem pretty simple, at least, simple compared to other numbers like the square root of 2, or π, or 3.843234, but it turns out that *real* numbers are, in a sense, simpler than them. So, Gödel's theorem actually doesn't apply to the real number system. Every truth about real numbers can be proven from the axioms of real numbers.

And there are many other mathematical systems that are simpler than arithmetic. But there are also many mathematical systems that are more complex than arithmetic. Nobody yet knows what the correct mathematical system is to describe the Universe is, so it's currently an open question of whether Gödel's incompleteness theorem applies. But, if it does, then I think that would be very interesting. There would be things about the Universe that would be impossible to "explain," and so we'd have no hope at arriving at a true "theory of everything."

Would this necessarily be a bad thing? Well, it might be depressing at first to discover that we can't know everything, but what's more important — the answer or the journey? If Gödel's theorem applies, then this would mean that physicists would always have a job! There would always be new

[3] However, apparently there *do* exist people who believe in the existence of "true" contradictions. This view is known as "dialetheism." I can see forms of dialetheism, like *psychological* dialetheism, making sense — often people hold conflicting, sometimes contradictory beliefs — but I have a hard time making sense of *logical* or *physical* dialetheism.

But who knows? Perhaps there is a meaningful way for true contradictions to manifest themselves in Nature. I think it's always important to keep an open mind. Being skeptical is, of course, also crucial to making sure we're thinking straight, but I feel we should always try to be open-minded skeptics when we can.

truths about the Universe to discover, and the process of discovery would never end. Whether this is a good thing is up to you.

Of course, even if Godel's theorem *doesn't* apply to the mathematical system that describes the Universe, we still wouldn't have explained *everything*. After all, we'd still be left with the task of explaining *why* that mathematical system is the one that describes Nature. Conceivably, the Universe *could* have been different, so why those axioms and not others? I turn to this very important question in *Lecture Notes 8*.

2 Types of Existence

Now for a brief digression. Up to this point, I've been rather casual in referring to the "theory of everything." In *Lecture Notes 6* I said "there's matter, energy, and forces. That's about it; that's everything." But what *is* matter, what *is* energy, and what *are* forces? Scientists generally tend to shy away from questions like these, claiming they're not scientific but are more philosophical than anything else (which is true). And they'll say "Oh, we don't concern ourselves with questions like that. We only care about scientific matters."

But, frankly, this kind of response confuses me. How can you feel comfortable *doing* science if you don't actually know what science is *about*? I personally would like to know what an electron actually *is* and what gravity actually *is*. I'd feel a lot more comfortable knowing what these things are before I have theories about them ... even if it means doing (gasp!) philosophy.

This leads us to a very big question, perhaps the Ultimate Big Question:

What is reality?

We find ourselves in a world where there exist many things, like matter, energy, and forces. What is the nature of the existence of these things? And do there "exist" in some sense other *kinds* of things? What is real?

Let me first explain what I mean by "nature of existence." There are at least a few "kinds" of existence that people have in mind when they talk about this sort of thing. The first is that of the stuff that the theory of everything is supposed to describe — *physical* existence. I don't aim to *explain* the nature of physical existence here; I just want to give the type of existence of matter, energy, and forces a *name* — "physical" existence.

The second kind of existence people talk about is *mathematical* existence. As I mentioned earlier, although we may seem to see math around us in our everyday life, like numbers and various shapes, there are many aspects of mathematics that we *don't* see. In fact, mathematicians often like to pride themselves in studying objects that have absolutely nothing whatsoever to do with the physical world. They like to study abstract ideas — constructed from certain axioms, of course — but which don't necessarily need to bear any similarity to the physical world. So, in addition to the physical world, there seems to be another kind of world: the mathematical world.

Finally, there seems to be a third kind of existence — *mental* existence. We have thoughts and perceptions about the physical world and about the mathematical world, but our conscious experience seems to be, at least on the face of it, different from something physical or mathematical.

These three kinds of existence, although they seem different at first, may actually be related to each other in a variety of ways. To begin with, most scientists today would say they believe that mental existence actually *emerges* from physical existence. There exist physical systems called humans, and inside humans is a vastly complicated network of things we call organs, cells, neurons, and so forth. Now, one of these physical things inside humans is called the "brain," and it is the control center of the human, telling him or her what to do. But, not only is it responsible for behavior; it's also responsible for thoughts, emotions, and the whole conscious experience. There are many things about consciousness that cognitive scientists still don't understand, but most believe that ultimately we will have a physical basis to mental existence.

I should say that there are some people (mostly philosophers) who don't share this viewpoint but who believe that the mind and the brain are actually separate things; these people take a "dualist" perspective of the mind. But I'm not going to argue for either case. It's a vastly complicated issue, and I don't aim to get into the so-called "mind-body" problem here.

Another connection between the types of existence that people have noted is the connection between mental existence and mathematical existence. This is the question of whether math is *invented* or *discovered*. Is math simply the product of the human mind that we create with their thoughts? Or is it something that exists "out there" independent of human or any physical existence?

This is another subject of debate, although it is probably less asymmetric in opinion than the mind-body debate. As with the mind-body debate, I won't try to give an argument here for either side, but both views have their pros and cons. Note that if math is invented, then mathematical existence simply arises from mental existence. And so, if mental existence arises from *physical* existence, then math really arises from physical existence. And, so, if this is the case, then the answer to the question of "what is reality?" is just "well, reality is this thing, and we give this thing the name 'physical.' It consists of things we call matter, energy, and forces, and everything else that exists in reality emerges from these fundamental physical things. In particular, math and mind emerge from the physical."

And this is certainly a valid viewpoint you can take concerning the nature of reality. But you're still left with a big unanswered question: if mathematical existence emerges from physical existence, then why does math seem to be so darned effective at describing reality? And, if mathematical existence *doesn't* emerge from physical existence, the question seems to be even more pressing: why should this abstract thing that exists *out there* describe *physical* things so well?

If you can't think of a good answer, don't worry. Most people in history who have thought about this have been quite puzzled by the "unreasonable" effectiveness of mathematics. So you have a choice: you can either accept this puzzling fact, or you can try to explain it.

Because I like the idea of trying to explain things until you are seriously stumped, I'd now like to present a very interesting explanation for this puzzle that actually *denies* the fact that math arises from matter. Briefly, the argument is that the Universe *is*, in a certain precise sense, mathematical, and therefore it's no surprise that math should be effective at describing the world! It's called the Mathematical Universe Hypothesis, and was proposed by Max Tegmark, a cosmologist here at MIT.

3 Mathematical Universe Hypothesis

3.1 The External Reality Hypothesis

We begin by making the philosophically radical claim that our mental existence isn't the only type of existence, *i.e.*, that there exist things besides my mind and your minds. OK, it doesn't sound so radical at first, but it does become quite a liberal stance once you think about it.

Strictly speaking, the only thing which we know for *sure* exists is ourselves

— well, the only thing *I* know for sure exists is *me*. *"I think, therefore I am."* It could very well be that all of these perceptions I have are nothing more than that. It may be that these perceptions are not reflections of things that exist outside my mind but are really just things that exist on their own. I exist, and everything I perceive is a product of my mind.

This view of the world is called "solipsism." It's a very interesting view and is actually extremely conservative. I certainly exist, but it's quite a leap to say that *you*, as well as a vast "external reality," exists. We're certainly used to making this leap but, when you think about it, it really is a bold philosophical move.

Nonetheless, I should say here that almost *nobody* in the world believes in solipsism, although you're perfectly free to believe in it if you want; I think it's an absolutely valid philosophical position to take, when formulated precisely. I *personally* (and many other people personally) think it's unlikely to be true, however. Why? Because a solipsistic theory seems to be necessarily more complicated than a non-solipsistic theory. By Occam's razor, then, solipsism is likely to be wrong. (Occam's razor, described in the Appendix of these notes, is the "rule" that the simpler the theory is, the more likely it is of being true.)

For example, even if solipsism is true, you don't really have control of the world. Things happen in your mental experience that are beyond your control; you can say they happen in the "unconscious." So when we do scientific experiments, for example, what we're really studying is our unconscious. But then it seems that the word "unconscious" simply becomes a label. In non-solipsistic theories, we have the conscious mind and the external world; here we have the conscious mind and the unconscious mind. So it seems the distinction between the unconscious mind in solipsism and the external world in non-solipsism is just a *naming* difference.

But there is a difference between the conscious mind in solipsism and the conscious mind in non-solipsism. I suppose there are many types of non-solipsism, but in the type of non-solipsism where the conscious mind emerges from physical reality — an external world — the conscious mind is just that: a physical thing that emerges when you have a complex enough system, like the brain. Consciousness in the physical world, then, is simply a consequence of the laws of physics.

However, consciousness in solipsism seems to be more complicated. Presumably your conscious mind operates under some rules, and presumably there are some rules governing how the conscious mind interacts with the unconscious mind. So it seems you've got 3 things to explain: (1) how does the conscious mind work, (2) how does the unconscious mind work, and (3) how do the conscious and the unconscious mind interact? In non-solipsism, conceivably everything could simply due to physics — that's all. In solipsism, you've got more explaining to do.

You might say, "well, maybe there are some *overarching rules* that govern everything in solipsism — how the conscious mind works, how the unconscious mind works, and how the conscious and the unconscious interact." But if you do that, then you might as well call those rules "physics"! The distinction between mental and physical reality, then, merely amounts to a labeling difference again, and you're back to non-solipsism. However, you could keep true solipsism without having these overarching rules, but then you're faced with giving the necessary 3-fold explanation I just described.

So that's why I don't favor solipsism — solipsism seems more complicated than non-solipsism and is therefore unfavored by Occam's razor. But you're perfectly free to believe in solipsism, and I'd be very curious to hear why you believe in solipsism if you do. I think it's a very interesting idea, and I think it'd be very interesting if it's true. (But these are merely emotions of mine, and why should reality care about human emotions?)

Back to the mathematical universe hypothesis (MUH). Tegmark arrives at the MUH by doing the (I think) very reasonable thing of rejecting solipsism and accepting the "external reality hypothesis," which is simply that an external world — an external reality — exists.

Now, if an external reality exists, then that reality should be describable in a language that isn't specific to us humans. For example, if intelligent aliens exist, then they too will uncover secrets about the external reality; and, if artificial intelligence can exist, then secrets about the external reality will also be accessible to AI. Now, we humans have a very particular way of describing that external reality; we have a certain set of concepts that we use in describing it. For example, we have notions like "protons" and "energy" and "the quantum state." Aliens might use very different concepts in describing reality — who knows how alien psychology might work?

Nonetheless, there will be similarities between a human description and an alien description. In both cases, we're trying to explain the external world, which consists of many things that have certain relations between them. But a true theory of everything shouldn't include any human-specific or alien-specific language, *i.e.*, "baggage." It should consist entirely of the things that the Universe consists of as well as the way that the things are related to each other. And, since it's free of any "baggage," it should therefore be something that is purely abstract.

Now, a set of abstract objects with relations between them is precisely the definition of what's called a "mathematical structure." So the external reality hypothesis *implies* that the external reality is a mathematical structure. So we arrive at the mathematical universe hypothesis: the Universe is a mathematical structure.

We therefore have a very simple explanation to the question of why math is so good at describing reality. The answer, according to the mathematical universe hypothesis, is that reality simply *is* mathematical.

I'll have more to say about this idea next week, as well as the arguably even crazier idea that our universe is a computer simulation.

4 Appendix: Occam's Razor

When scientists come up with theories to describe the Universe, they generally try to come up with the *simplest* theory they can — *the simpler the theory, the better*. This "principle" is known as *Occam's razor*, after the philosopher who first used it about 700 years ago.

It's a reasonable-sounding little principle, on the face of it, but why should it be true? Well, first of all, we need to be clear what we mean by the "best" theory. Obviously, what we have in mind is "the theory that is most likely to be true." We want to understand Nature, and we want the theory that has the greatest chance of being right. I'll define more precisely what we mean by a "simple" theory shortly (and I'll let you see if it conforms to any intuitive ideas of simplicity you may have).

Every theory that we've ever used to describe Nature has some fundamental principles or *"postulates"* — these are the fundamental, unprovable statements about Nature on which the entire theory is based and from which all consequences of the theory are drawn. For example, remember that special relativity had 2 postulates: (1) the laws of physics must be the same for all observers moving at constant velocity, and (2) the speed of light must be the same for all observers moving at constant velocity. Other theories are built upon other postulates.

Unfortunately, it's impossible to know if the fundamental postulates of a theory are true.[4] You can certainly perform many experiments to *test* those

[4] At any rate, it definitely *seems* impossible, and I've never heard of a way to get

postulates — the postulates of special relativity have been consistent with all experiments ever done, for example — but there's still the chance that they're wrong, or at least *incomplete* in some way. It could be that the true nature of Nature is a bit deeper than the postulates of relativity (more precisely, general relativity), and we've just been *fooled* for all of these years.

For example, for a long time it seemed that Newton's law of gravity (which may be taken as a postulate of the Newtonian "theory of everything") was an accurate description of Nature, because all observations at the time were consistent with it. Eventually, however, observations were made which *contradicted* Newton's law of gravity (a famous example being the anomalous "precession" of Mercury's orbit). This meant that the postulate of Newtonian gravity wasn't completely right, and needed to be replaced (the replacement being the postulates of general relativity, which accurately predicted the precession of Mercury, for example). So, Newton's law of gravity isn't *entirely* wrong — it very accurately describes the gravitational effects of a wide range of phenomena. It's just, in a sense, "incomplete," because it is inaccurate in accounting for the gravitational effects in many other phenomena. (Note: this kind of "incompleteness" should not be confused with "incompleteness" in the Gödel sense!)

So, there's always some uncertainty in the postulates of a theory. Now, the best justification of Occam's razor that I've heard goes as follows. Let's *assume* that every postulate of every physical theory has a fixed amount of uncertainty; *i.e.*, every postulate has the same amount of uncertainty associated with it as every other postulate. Then, the more postulates you have in a theory, the more uncertainty you have in that theory. We want to minimize uncertainty in a theory, so we should minimize the number of postulates in the theory. If we define a "simple" theory as a theory with a small number of postulates, then it becomes clear that the simplest theory has the least amount of uncertainty, and is therefore the one that has the greatest chance of being right. The argument obviously isn't airtight, but it's the best I've heard.

---

around it. I only include this footnote because I'm extremely hesitant to ever call anything "impossible." It could be that we're simply not *smart* enough or *lucky* enough right now to have come up with a way of proving that the axioms of a certain theory are the axioms of the true theory of Nature. I am reminded here of Arthur C. Clarke's first "law" of prediction: *When a distinguished but elderly scientist states that something is possible, he is almost certainly right. When he states that something is impossible, he is very probably wrong.*