Musements

Explorations in Anthropic Reasoning

Karl Stiefvater

Contents

Counting

Suppose I asked you to draw a card from a deck. The deck, I tell you, has cards numbered from one to N, where N is a number I know, but I'm not telling you.

Suppose you draw the card numbered four. Can you make any guesses as to the value of N?

You can be sure that N is not one, two, or three. Four is an okay value for N. So is five, and any number above. But 10,000,000 is probably a bad guess for N – if N were so large, you'd expect to have drawn a number much larger than four.

Now, let's change things around.

This time, I tell you that N is the number of people who will ever exist on the planet earth. Suppose I tell you you're the six billionth person. Can you make any guesses as to the value of N?

Intelligence

Okay. So John Searle doesn't believe that it is possible for machines to become intelligent. One of his arguments is the "simulation isn't real" argument. It goes something like this. I think.

First Searle acknowledges that computers can simulate physical systems. Like molecules making a gas. Or planets rotating about each other. Or the weather. Or the neurons in the brain of a human being.

BUT he says – this is merely a simulation. Like when a computer simulates a plane flying through the air. Sure, he says, inside the computer the plane is flying – but it can never, ever, carry passengers from New York to San Francisco. That's the difference between a simulation and the real deal. And the same is true when a computer simulates a human's brain. It may act like a human – but it's not the real deal. It has no real intelligence.

I disagree. Consider.

What if we dig-up Beethoven. We get his genetic code. We get the genetic code of everyone alive during his life. We get the plans for all the buildings, all the forests, EVERYTHING on the planet earth the day that Beethoven was born. And we simulate it.

(Yes, this is a very nearly impossible task. Certainly entirely impractical. But we're conducting a thought experiment, so such things are allowed.)

Now we simulate the entire earth, from the moment of Beethoven's birth. Is the simulated Beethoven intelligent? (Nevermind the other simulated people living with him.)

Searle says no. Yes, he looks like Beethoven (inside the computer he does, anyway.) And he acts like Beethoven. And he composes like Beethoven. But no – the real Beethoven existed long ago – this doppelganger is merely walking in the original's footsteps. Nothing original. Nothing truly intelligent.

Um.

Well, it is a simulation after all. What if I pause it for a moment, step inside with my digital scalpel, and cure the simulated Beethoven of his simulated syphilis.

Now – when the simulated Beethoven finishes his 9th symphony – he doesn't die. He (and the rest of his slightly modified world) continue on. He writes his 10th symphony.

I want to know why this 10th symphony isn't an intelligent work.

I would love to hear it.

Simulations

Have you seen that movie "The Matrix"?

So – I don't know how to tell you.

It's a true story.

Scary, huh? The guys running our simulation probably aren't evil robots. After all, if they wanted our energy wouldn't they just burn us?

But yeah, we live in a simulator.

How do I know? Well – it's a pretty simple argument:

Let's assume that someday someone will invent a perfect simulator like the Matrix. Doesn't matter when – might be in 10 years, might be in 10,000. Just has to happen someday.

Given how humans reproduce exponentially, our technology advances exponentially, and all – it's safe to assume that sometime after the perfect simulator is invented, there will be few of them, then dozens, then thousands. Millions. Billions, Bajillions –

given enough time. Again, it doesn't matter when, it just has to happen someday.

And here is the key point: if you can't tell whether you're in the real universe or in one of the future perfect simulator universes, you have to accept that maybe you DO live in a future simulated universe.

So you can compute the probability that you're in a simulated universe. Simply count the simulated universes vs. the real universes. The ratio tells you your chances of being in one or the other.

Universes

So, a few years back I had an idea. A great idea, or so I thought. The best idea of my life – the best idea I'll ever have.

Like all good ideas, someone else had it first.

But I'm over that now. And since I still think it's a good idea, I ought to spread the word.

I've discovered where the universe came from.

Obviously, if you're an atheist, this is an important question. But even theists have a problem here: if god created the universe, where did god come from?

It's a nasty problem. Maybe it was Saint Thomas Aquinas who first found it. But I've got a handle on it now.

Consider the number three. Not the digit, nor the symbol, but the number itself. The thing that comes after two.

Where did it come from? Who made it?

If you're like most people, you'll say "It didn't come from anywhere – it doesn't really exist. It's just a concept. Maybe somebody discovered it, but it wasn't really *made*."

Mathematicians have a name for this: mathematical existence. Unlike physical existence, mathematical existence isn't *real*. You can't slam your fist against the number three. Lots of things exist mathematically (strange multidimensional singularity fractals, or somesuch) but nobody expects that they really exist.

So how about the number π? That's the one that goes $3.141\ldots$ on and on forever. It's another number. It has fantastic properties, which is why mathematicians gave it such a good name.

But – who made it?

"Nobody made it, because it doesn't really exist."

Right. In fact, whenever there's a new mathematical discovery, that's what it's called, a *discovery*. Did Mandelbrot make the Mandelbrot set? No – he discovered it.

So – what about the people who live inside π? Who made them?

"Excuse me?!"

Okay. Maybe there aren't people living inside of π. But there ARE numbers that have people living inside of them – as you look down the row of digits, if you interpret them right, you see people's lives being played out. And I bet that people live in π, too – but that's not important right now.

What about them?

(Okay. Mathematicians are going to be pissy. "What numbers?" they say. Consider the rules for John Conway's game of life. It's been shown to be Turing complete – i.e. computation can be performed in his worlds. Create a Turing machine which executes a genetic algorithm – encode it into a life world – and then encode it into a number. There.)

So, what about them?

“They don’t really exist. They exist mathematically, as you say – but they don’t exist physically.”

No – I think they do.

Consider for a moment the perspective of one of these people living in one of these worlds. To you his world is intangible. But to him his world is real. When he slams his fist against his world, it hurts.

That is EXACTLY the same criteria that we use to judge our world to be real.

And consider that he can find a number that contains our world. And he says, just like you do, that our number-world doesn’t really exist.

“So, according to you, do we exist?”

No, that’s just it. Physical existence is an illusion. We exist mathematically. Our universe is merely a definition – like the number three.

Granted – our universe is quite big. But then, so is π.

If you've been following closely, you'll noticed that a *theory of everything* falls neatly from this package.

But let's approach from a different direction, shall we?

A theory of everything is the holy grail of physics. Back in Einstein's day, it was called a grand unified field theory, and it's what he worked toward most of his life.

Basically, it's the set of rules that describes why the universe is the way it is. It should explain EVERYTHING – once you have the theory of everything (TOE) you don't need any other theory. Nowadays we have several competing theories (relativity, quantum electrodynamics, gravity, string theory, etc.) not all of which play nicely together. The TOE will fix that.

Physicists and philosophers agree that the TOE – whatever it turns out to be – ought to be *elegant*. It should be simple, self evident, easy to state. It should offer insight into why things are the way they are.

Most physical theories these days are *not* elegant (no offense.)

The theory of gravity, for instance, says that masses attract one another – but offers no explanation why they attract at the rate they do – nor why they attract at all – why not repel?

And what about the big questions: why do we see three dimensions? Why only one dimension of time? What the hell is time? Why is the universe non-deterministic? Is it infinite? If not, why is it so enormously outrageously big?

And perhaps the most important question of all: why does the universe follow this particular set of rules, and not some other set of rules?

...

There is, I think, only one possible TOE which satisfies the requirement of elegance: they all exist. For each and every different possible TOE, there is a real live universe that follows its rules. The whole universe is, in fact, this collection of universes. The multiverse.

...

But this is exactly what we learned from our discussion of mathematical existence. In it, we said that anything which exists in the world of mathematics exists physically, too.

What are these mathematical things? They are anything which has a consistent ruleset to define it. A TOE if you will. And so they all physically exist. And the entire collection of them is the multiverse.

Well – it's time for a definition.

I'm a big fan of Turing; and mathematically I lean constructivist. So I'll take a risk and restrict myself to computable structures. Since we started all this with people living inside numbers, let's stick to numbers. (No worries – computable numbers are isomorphic to computable structures. And heck – Turing's big paper was about these numbers. So.)

Definition: the multiverse is the set of computable numbers.

...

For the layman – what does this mean?

A computable structure is anything that is produced in a step-by-step manner, following a fixed set of rules, where no intelligence is required. It's anything that could be produced by a machine.

Computable numbers are made in the same way. π is a good example – there is a well known ruleset for writing out the digits of π. It requires a lot of work – but the process is simple.

It may not seem like it, but we don't lose anything by considering only numbers. The more complicated structures (words, sets, hierarchies, arrangements of atoms, ...) can be encoded as big numbers. (Trust me, it's okay.)

Besides – the other universes are bound to be extremely strange to us. Maybe they don't have movement, only growth? We'd better convert everything over to numbers so that we're all on equal footing.

"Sure sure, this all sounds fine and good. But obviously we can never interact with the other universes, we can never go there, there's no way to test whether they really exist, now is there? So this is all a bunch of useless mumbo jumbo."

Au contraire.

If they do in fact exist, and if in fact we happen to live in one of them, then it stands to reason that the one we live in ought to be *typical*.

So what we do is look for common traits amongst the universes. And if we find that our universe doesn't have any of these traits, well then, heck, it's probably not one of them and they probably don't exist.

But if it does…? Then the theory has some merit. AND we've found an explanation for why our universe has that trait.

For example:

1. Why is our universe non-deterministic?

A typical universe in my system has a finite, but hugely enormous, Kolmogorov complexity.

What's that, you ask?

Suppose I ask you to pick a random number. And you say "What, between 1 and 10?" and I say "No, any random number."

So you say "9747". And I say, "Oh please! That's hardly random, it's so small!"

And you say "Okay, 97478346586342856328472323456."

And I say, "Ha! Can't you get any more random than that? It's so tiny."

For a good random number, with no upper bound, you need to pick one that's as big as the universe.

Aha!

(In our case you didn't pick a random number, instead you were born into a random universe.)

Here we find a good explanation for why the universe would be finite but so outrageously big. (Is the universe big? Count the particles and realize that each new telescope increases this number by magnitudes.)

More subtle, though, is the fact that the ruleset for a typical universe ought to have a similarly large set of rules. This is contrary to common intuition among physicists, but I think deserves closer examination.

According to quantum physics, our universe is constantly making quantum choices which are not strictly determined by its rules.

But machines can't make random choices. So if our universe is produced by a machine, all those choices must be written into the rules themselves. Now we realize our universe *does* in fact have a nearly infinite ruleset (i.e. enormous Kolmogorov complexity.)

And from this perspective, my collection of universes and the collection of worlds predicted by the many world interpretation (MWI) are very very similar.

2. Why doesn't the universe change its rules?

Here is a problem. Gravity never reverses itself, but it easily could. The programming of our universe could include the rule "on July 4 2069, gravity goes backwards." And certainly, in some universes, this is exactly what happens.

But this sort of behavior better not be the norm for universes, as I've never seen it happen here. Why?

I see two possibilities.

One, maybe our universe is somewhat special: it does have life after all. For life to have evolved, its rules needed to remain stable for a billion years, and perhaps there's some momentum that keeps our universe obeying them today.

Two – and this is my hope – perhaps rule-stability is a natural property of typical universes.

But I don't know; it's what I think about these days. If you have any ideas, let me know.

Immortality

It turns out we're immortal. It's a result of quantum physics.

Now – DON'T GET ME WRONG. I'm not talking about some new-age-quantum-physics-ufo-bigfoot immortality here. I'm talking about $2 + 2 = 4$, so we're immortal. I think it's really very freaky.

First of all, you need to believe in the many worlds interpretation (MWI) – more commonly known as "the theory of parallel universes."

Many people don't. But you should, really. It provides the only reasonable explanation for quantum experiments – like the two slits experiment. Before MWI, we had to say nonsense like "photons are *both* waves *and* particles." what the hell does that mean? It's one or the other, bub.

And anyway, nowadays people are building computers to take advantage of the parallel universes. And they work. So if you don't believe in the MWI – you'd better find some explanation for why these computers are computing 128 times faster than they ought to.

Besides – Stephen Hawking believes in the MWI. Don't you trust him?

But we're not talking about the MWI. We're talking about immortality.

Once you believe the MWI – it's easy to see that you're immortal. In some of the universes – very few – of the universes, you magically (luckily) survive all the nasty things that may happen to you. You don't get cancer. You don't walk down that dark alley. You don't get too close to that cliff.

And furthermore, in some of them, you don't seem to age as quickly. And in some of them, medical science makes discoveries 100 times faster than normal.

You get the idea.

Now – the clever thinkers among you will say "Yeah, yeah, big deal. In some of the universes I am immortal, but in overwhelmingly most of them, I'm not. Chances are I'm in one where I die."

You're right, you're right. Being immortal in some small number of universes isn't terribly interesting if there's virtually no chance of actually experiencing it. (You think it's hard to win the lotto – try winning a universe in which you're immortal!)

Here's where the argument takes a subtle turn. And I must confess, I'm not terribly good at conveying it.

I think (and others agree – we've all discovered this argument independently) you *must* experience the universe of immortality. For the simple reason: it's impossible to experience a universe where you're dead. Because there you're dead.

Here's a thought experiment that helps to think about it. It's not mine – I've stolen it from someone else. When I can remember who it was, I'll give him credit.

It goes like this: imagine you have a room with 10 computers – each of the computers is artificially intelligent. Each of the computers is running exactly the same program, and they all get exactly the same inputs, and none of the computers can talk to any of the other computers. So, in fact, they all do exactly the same thing, all the time. It's kinda silly, actually, because they're all identical – it's a waste – only one computer is necessary.

So, in fact, despite the fact that there are 10 copies– there is only one intelligence. One mind.

And if we switch off one of the computers, the mind doesn't notice.

Get it?

Epilogue

I wrote these essays during a curious period.

I got it into my head that I wanted to live in an RV on the streets of San Francisco. I wanted to live the life of a starving, street dwelling artist. I discovered that hunger is an extremely motivating force.

But since my life was in an unconventional place, and since these ideas come from a very unconventional place, my friends worried that I might be coming a bit unhinged. Seeing the reasonableness of their concerns, I too began to worry I might be coming a bit unhinged.

Happily, I began to discover that other people (much smarter and much saner than I) were also developing these ideas. So if I was in fact becoming unhinged, at least I was in good company.

The ideas explored in the first essay were first proposed by astrophysicist Brandon Carter in 1983. It has been repeatedly discovered and discussed over the years, and is now known as the "Doomsday Argument."

The second essay explores ideas in strong AI. My favorite text on these subjects is "The Mind's I" by Douglas Hofstadter and Daniel Dennett.

The ideas of the third essay have appeared many times in history – but perhaps the best exposition on them is that by Nick Bostrom, called "The Simulation Argument".

As for the fourth essay: the idea that mathematical and physical existence are one and the same can be traced back as far as Plato. However, the arguably best modern essay on the matter is Max Tegmark's paper "Parallel Universes".

First person immortality, as a result of parallel universes, is probably the most controversial idea here discussed – and reputable thinkers have smartly distanced themselves from the concept. Nonetheless, you can find it referenced in Max Tegmark's "Quantum Suicide Experiment", and also in essays by Hans Moravec and Bruno Marchal. Friends of Hugh Everett III (the father of the many worlds interpretation) claim he also discovered the argument, and his daughter cited the theory in her suicide note.

Lastly – and most surprising to me – is that each and every one of these ideas weave together to create the story of Greg Egan's book "Permutation City". It's truly a modern masterpiece – you simply MUST read it if you haven't already.

I hope I've piqued your interest.

Best wishes,

K.