Dwarkesh Podcast #83 - Adam Brown – How Future Civilizations Could Change The Laws of Physics

Published - December 27, 2024

Transcribed by - thepodtranscripts.com

Dwarkesh Patel
Today I'm chatting with Adam Brown, who is a founder and lead of the Blueshift team, which is cracking math and reasoning at Google DeepMind, and a theoretical physicist at Stanford. Adam, welcome.

Adam Brown
Delighted to be here. Let's do this.

Dwarkesh Patel
Okay, we'll talk about AI in a second, but first, let's talk about physics. First question: What is going to be the ultimate fate of the universe, and how much confidence should we have?

Adam Brown
The ultimate fate is a really long time in the future, so you probably shouldn't be that confident about the answer to that question. In fact, our idea of the answer to what the ultimate fate is has changed a lot in the last hundred years. About 100 years ago, we thought that the universe was just static, wasn't growing or shrinking, was just sitting there statically. And then in the late 20s, Hubble and friends looked up at massive telescopes in the sky and noticed that distant galaxies were moving away from us and the universe is expanding.

So that's like big telescope discovery number one. There was then a learned debate for many years about, you know, the universe is expanding, but is it expanding sufficiently slowly that it'll then recollapse in a big crunch, like a time reverse of the Big Bang, and that'll be super bad for us? Or is it going to keep expanding forever, but just sort of ever more slowly as gravity pulls it back, but it's fast enough that it keeps expanding?

And there was a big debate around this question, and it turns out the answer to that question is neither. Neither of them is correct. In possibly the worst day in human history, sometime in the 1990s, we discovered that in fact, not only is the universe expanding, it's expanding faster and faster and faster.

It's what we call dark energy, or the cosmological constant. This is just a word for uncertainty. Is making the universe expand at an ever faster rate, accelerated expansion as the universe grows?

So that's a radical change in our understanding of the fate of the universe, and if true, is super duper bad news. It's really bad news because the accelerated expansion of the universe is dragging away from us lots of distant galaxies. And we really want to use those galaxies.

We have big plans to go and grab them and turn them into vacation destinations or computronium or in any other ways extract utility from them. And we can't, if the cosmological constant is really constant, if this picture is correct, because anything close enough, we can go out and grab it, obviously. But if it's further away than about a dozen billion light years, the expansion of the universe is dragging it away sufficiently rapidly that even if we send probes out at almost the speed of light, they will never make it.

They will never make it there and make it back. They'll never even make it there if it's sufficiently far away. And that means that there's a finite amount of free energy in our future.

And that's bad. I mean, that means we're doomed to a heat death if that's true. But is it true?

I mean, that was the second half of your question. And first of all, we keep changing our minds about these things over the last century or so. So on first principles grounds, you may be somewhat suspicious that we'll change our minds again.

And none of this is settled physics. And indeed, it may be that the cosmological constant is not constant and you should hope with all your heart that it's not. It may be that it naturally bleeds away.

It may be, in fact, that our fate is in our hands and that our distant descendants will go and bleed the cosmological content away, will force it to go to zero. They will be strongly incentivized to do it if they can, because otherwise we're doomed to a heat death.

Dwarkesh Patel
How would they bleed this away?

Adam Brown
This obviously depends on physics that we're not totally sure about yet, but it seems pretty consistent with the known laws of physics that the cosmological constant, what we perceive it as being a constant, this dark energy quantity that's pushing the universe apart from each other in many very natural extensions of the known laws of physics. That is something that we have the ability to change. In fact, it can change, can take different values.

It is not just totally fixed once and for all that in fact, you have what's called a different vacuum, different regions of parameter space that you can transition between, in which the cosmological constant can take different values. And if that's true, then, well, you can either sort of wait around and hope to get lucky, hope that the universe just sort of spontaneously moves from one of these vacuums to another one with a lower cosmological constant tending towards zero asymptotically, or you could take matters into your own hand, or you could imagine our descendants deciding that they're not going to just suffer the heat death, that they're going to try and trigger a vacuum decay event to get us from the vacuum we're in to another vacuum with a lower cosmological constant, and our distance descendants will be forced basically to do that if they don't want to suffer a heat death.

Dwarkesh Patel
Yeah, Proceed with caution, but definitely proceed.

Adam Brown
With caution in these theories where there's lots and lots of vacuums out there. And most of those vacuums are incredibly inhospitable to life as we know it. In fact, seemingly they're just completely inhospitable to all forms of intelligence.

So you really, really don't want to end up in them. However, again, if our best theories are correct, it seems as though there should be some of them that are much like our own in many ways, but have a lower value of the cosmological constant. And so what we'd want to do is engineer that we end up in one of those vacuums.

Dwarkesh Patel
Sorry, what is a vacuum?

Adam Brown
Ah, great question. A vacuum is like a possible, well, what we would perceive as a possible set of laws of physics as we see them. So what it really is, is a minima in some higher dimensional abstract laws of physics, space in which you can find yourself in a minima.

But these minima may just be local minima. In fact, according to our understanding, the minima which we live today, that gives us all of the laws of physics that we see around us, is in fact just a local minimum. And there's a lower minimum.

In fact, there's many lower minima out there to which we can transition spontaneously or because of our own deliberate action.

Dwarkesh Patel
Okay, I'm just going to throw all my confusion at you and you figure out which one is worth dealing with first. What is the nature of the loss function that makes one value a minimum and one higher up? You know, what is exactly the ball rolling up on when it gets out or into a valley here?

And then you're hinting at the possibility that there are other places in. I'm not sure if you're suggesting in the physical universe or in some hypothetical universe where the vacuum could be different, as in, in reality there are other pockets with different vacuums, or that hypothetically they could exist, or that our universe counterfactually could have been one of these? I don't know.

This is the kind of thing I'd like throw into. Put everything I can into a Claude prompt and see what comes out the other end.

Adam Brown
Good. Well, I'm happy to be your Claude. The loss function is the energy density, and so maybe a good analogy would be water.

Water can exist in many phases. It can be steam, it can be water, it can be ice. And even if it's in a cloud, let's say it would rather be water than be water vapor, but it's having a tough time getting there because in the middle there's a barrier.

And so, you know, spontaneously, it can eventually, due to a sort of thermal process, turn from steam into water. These will be like the two minima in this loss landscape. Or you can go and do cloud seeding to turn it from water vapor into water, and so those would be the equivalent of the minima here.

The existence of different minima in general is a very well-established part of physics. The possibility that we could engineer going from one minimum to another in a controlled way is a more speculative branch of physics speculation, but it seems totally consistent with everything we know that our distant descendants would try to attempt it.

Dwarkesh Patel
What would it take to do this?

Adam Brown
Probably you'd want something that would look a bit like a particle accelerator, but it would be considerably more controlled. You'd need a very controlled way to collapse a field and make a bubble of this new vacuum that was big enough that it would continue to expand, rather than just recollapse under its own surface tension.

You would have to do that in a very careful way, both to make sure that you didn't accidentally make a black hole instead by the time you've concentrated all those energies, and also worse than making a black hole, would be ending up in a vacuum that you didn't want to end up in. It would be ending up in a vacuum in which you had not only bled off the cosmological constant in some way, but that you had changed, let's say, the electromagnetic constant or the strong nuclear force or any of these other forces, which would be seriously bad news.

Because if you did that, your life, as you know it, is extremely well attuned to the value of the electromagnetic constant in your evolutionary environment. It will be very, very bad indeed if we changed those constants as well. We'd really just try and target the cosmological constant and nothing else, and that would require a lot of engineering prowess.

Dwarkesh Patel
So, it sounds like you're saying that changing the laws of physics is not even Dyson sphere level crazy. Somebody could do it on some planet in the middle of nowhere.

Adam Brown
I'd say it's definitely substantially harder than Dyson spheres as far as the tech tree goes, but it's not magic. We're not actually changing the laws of physics. We're just changing the low energy laws of physics as they present to us in this scenario.

Again, this is speculative, but it's not super duper crazy. It's a natural consequence of our best theories, or at least some of our best theories of quantum gravity, that they allow for this possibility.

And there is a meta law of physics, the true laws of physics, be it string theory or whatever else, that you're not changing. That's just the rules of the game. What I'm describing is changing the way that the universe looks around you, changing the cosmological constant.

So I think again, changing water into water vapor into water is a great analogy here. There's nothing actually — the laws of physics are still the laws of physics, but the way it feels to live in that universe, the value of the electromagnetic constant is perhaps not an absolute fixed value. It can vary in different places.

Similarly, the density of water around you, the viscosity would change. It'll be an environmental variable like that.

Dwarkesh Patel
One question you might have is if this is the thing that could, maybe spontaneous is the right way to describe it, if this is the thing that can just kind of happen, there's something really interesting where if a thing can happen, you kind of see examples of it happening before. Even with nuclear weapons, I don't remember the exact phrase, but wasn't it the case that early in Earth's history, when there was a higher fraction of 238 isotopes, that there were spontaneous nuclear explosions?

Adam Brown
There probably was spontaneous nuclear reactors. They've discovered a seam in Africa where it looks like there was a fission reaction that naturally happened. It didn't explode, but it did do the same thing that happens in our nuclear power plants.

Dwarkesh Patel
One way you can look at nukes is like, this thing just would not have been possible if some intelligent beings hadn't tried to make it happen. But something like this happened before because the laws of physics allow it. Is there any story you can tell here where this vacuum decay, maybe it takes a super intelligent species to coordinate to make it happen, but also because it is the thing that the laws of physics can manufacture or can allow for, it has happened before or is happening?

Adam Brown
Yeah, absolutely. Almost certainly anything that humans can do can happen without humans. It's interesting to reflect on what aspects of human behavior nature has a tough time doing without us and what it just does on its own.

For example, we make colder things in our laboratories than really exist naturally in the universe. But the universe certainly could make anything colder just by chance.

Vacuum decay is something that if it is possible, will in our future definitely happen. That's just a feature of the world that eventually in our distant future, if it's possible at all, it will happen due to a quantum fluctuation.

Our descendants may not wish to wait around for a quantum fluctuation to happen. They may wish to take the fate into their own hands, since a quantum fluctuations can take exponentially long times to happen. And if they even happened, you'd end up in an unfavorable vacuum, not hospitable for life, rather than trying to steer the cosmological constant in a happy direction.

But they certainly can happen, and in our future, and indeed definitely will happen, if they're permitted. According to our understanding of quantum mechanics, if they're permitted, they must eventually happen.

Furthermore, there are again speculative, but not wild theories of the early universe in which this happened in our past, in which we transitioned far, far in the past, maybe into what's called a bubble universe. So we started off in some other much higher vacuum long in the past. And then what we see as the Big Bang was in fact just a sort of local vacuum decay that then gave rise to our the bubble in which we live, everything we see around us.

Dwarkesh Patel
Who would be in a position to seed these bubbles?

Adam Brown
Usually people are thinking that something just spontaneously happens, in the same way that rain spontaneously happens in a cloud, that somebody didn't go and seed it deliberately

to make it happen. But you could more than free to speculate that somebody seeded it to make it happen as well.

Dwarkesh Patel
How does this respect the conservation of energy?

Adam Brown
Energy, or the conservation of energy, is not conserved in general relativity. Energy is not conserved. It's conserved locally, at things you can do at a local level.

But in an expanding universe, energy is not conserved globally. This is one of the big surprises.

That is not a speculative statement. That is a statement that goes all the way back to Einstein and general relativity: energy is simply not conserved at the global level. It's conserved at the local level.

You can't do something in your lab that will generate free energy. But if you can participate in the expansion of the entire universe, then energy is not conserved.

Dwarkesh Patel
So if you were to spawn a bubble universe in your lab, you've theoretically created a lot more matter and energy. What would be the thing that offsets this or that makes this viable?

Adam Brown
Energy is conserved in a universe that's not expanding, a static universe. A universe that is expanding, energy is not conserved. It can just appear.

General relativity is quite clear on that. General relativity, Einstein's theory of space and time, one of our most beautiful and best tested theories, is quite clear on that point.

Energy is not conserved. To ask what happened to the energy, you can ask at a local level what happened to the energy density. But at a global level, energy is simply not conserved.

Dwarkesh Patel
Then do our future descendants have any constraints in terms of... Because earlier we were mentioning, as a catastrophe, we found out about the cosmological constant because it limits our cosmic horizon and thus limits the free energy that our descendants would have access to. But if you can just make entire universes...

Adam Brown
Then this is a matter of extreme interest, I would say, to us. It won't be relevant for tens of billions of years, probably, because that's the timescale on which the cosmological constant operates.

But if the cosmological constant is truly constant, and we've only known about it for 25 years, and there are astronomical observations that seem to be in tension with that, but if it is truly constant, then there is a finite amount of free energy in our universe. If it's not constant, if we can manipulate it, or even if it naturally decays on its own, then there is the possibility of an unbounded amount of free energy in our future, and we would avoid a heat death scenario.

Dwarkesh Patel
The situation you mentioned earlier, where somebody seeded our universe, they've created a bunch of energy. That's related to them having something equivalent to a positive cosmological constant in there.

Adam Brown
Yes. In any of these scenarios in which our universe is a bubble that formed in a sort of bigger, what's called a multiverse, or that's a loaded term, but a sort of larger universe in which our universe is just one bubble, the higher meta universe also has a cosmological constant, and it is higher than the value in our universe. That is the one sense in which there's some version of energy conservation: you can go down from high to low. It is considerably harder to go from low to high.

Dwarkesh Patel
So the idea is that you'd recursively have universes in which the bottommost one would immediately implode because of a negative cosmological constant, and the biggest one is exponentially increasing.

Adam Brown
Correct. The rate at which the universe is exponentially increasing is set by the cosmological constant, in which the volume of the universe is exponentially increasing. So you can imagine a scenario in which there was a high cosmological constant.

You have a bubble universe that has a lower value of the cosmological constant. It continues to expand.

You could make new bubble universes or new regions in that universe that have a lower cosmological constant, either naturally and spontaneously or due to action that we might take. And as long as that cosmological constant is non-negative, is zero or positive, that universe will not implode. If it goes negative, that universe will eventually implode.

So you could imagine a cascade in which you go to lower and lower values of the cosmological constant. There are a lot of engineering details to be worked out, but what I'm describing is a scenario that is not inconsistent with the known laws of physics.

Dwarkesh Patel
How likely do you think this is?

Adam Brown
If the laws of physics are as we believe them to be, and if we do not blow ourselves up in some other way, this is an issue that our distant descendants will eventually have to confront.

Dwarkesh Patel
No, no, no. As in the whole, like, there's like other bubbles. Not about something our descendants might do, but the fact that the big bang was the result of a bubble within some other metastable state.

Adam Brown
That's a tricky question. But since you asked it, I'd say probably 50%.

There's a lot we don't understand about any of these questions. They're all super speculative. It's an active area of research how to combine quantum mechanics and expanding universes.

On the other hand, it seems pretty natural when you do combine quantum mechanics and gravity and try and fit them all together in a consistent picture. If universes can expand a lot, then at all, according to the gravitational theory, then quantum mechanics will naturally populate those bits that can expand a lot, and so you'll naturally end up with an expanding universe. So I would say probably in my heart, slightly higher than 50%, but I'm going to round it down to 50 out of epistemic humility.

Dwarkesh Patel
It's funny because this is often the way people talk about their AI timelines of, like, really, I think it's like 2027, but if I'm taking the outside of you, I'm going to say 2030. Okay.

And is there any way, given our current understanding, of using bubble universes to do useful work for the people outside of it? So to do some computation within it or to get some sort of actual energy out of it for the people.

Adam Brown
Outside of the bubble? So the thing about these bubbles is that they tend to expand at the speed of light. So even if you start off outside, you're probably going to end up inside them in short order unless you run away very quickly.

So this isn't something that we make in the lab and then just remains in a box in the lab. And then we use use to do things. This would be something that we would do or maybe would just happen to us because of spontaneous vacuum decay and it would engulf all of our future light cone.

And so we wouldn't. It's not a it's not a box that you're using to do things. It's a new place that you live. You better hope that you've engineered the stuff so that that new place is still hospitable for life.

Dwarkesh Patel
So look, if it's the case that you can set up some apparatus, not now, but not in this room, but eventually, that if some individual wants to change the constants of nature, they can not only do this, but then the repercussions will extend literally as far as light can expand. You might have some hope that future civilizations, individuals or AIs have tons of freedom, so they can do all kinds of cool things. You can have your own galactic cluster over there, and if you want to go do whatever you want, right, go live your life, and there's going to be some libertarian utopia. But if you can literally destroy the universe, maybe it's a different story.

Adam Brown
That is a big negative externality, destroying your future light cone. And in a world with big negative externalities, libertarian fantasies can't really happen. It has pretty good big government governance implications, is that if it is possible for people just to wipe out their entire future light cone — not only themselves, but everybody else who wishes to participate in that future light cone — then we're going to need a government structure that prevents them from doing so.

The worst-case scenario is even worse than that. Not just that they could do it, but that they in some sense be incentivized to do it. You could imagine really adverse laws of physics in which maybe you could speculatively build some power plant that just really makes use of sitting on that edge of instability. And then each person individually might say, "Oh, I'm quite happy to bear one in a trillion chance that I wipe out the future light cone because I get so much benefit from this power plant."

But obviously the negative externality means that people really shouldn't do that. So I hope the laws of physics don't turn out that way, otherwise, we're going to have to have some super-arching control.

Dwarkesh Patel
I've done a couple of these interviews, and these end up being my favorite interviews, where a normal person who has just had grade school education can think, "Of course, I understand this, right?" Or if you've just seen enough YouTube videos about pop sci. Give you a concrete example. When I interviewed David Reich, the geneticist of ancient DNA, I feel like we have a sense that we understand the basics of how humans came to be.

What is the story of human evolution? And just like the episode revealed to me that the main questions you might have about how humans came to be — where did it happen? When did it happen? Who did it happen with? In fact, the last few decades of insights have totally revolutionized our understanding.

We have the sense that we understand what basically cosmology implies, but this idea that, in fact, there's this underlying field which not only implies very interesting things about the distant past, about the Big Bang, but also what our future descendants, you know, what kinds of civilizations they'll be able to set up, both from a governance and a practical energy perspective, it totally changes your understanding.

Adam Brown
It just keeps changing. Not just your idea, our idea, everybody's idea has changed a lot in my lifetime and may continue to change. In some sense, it's because you have the lever arm, the long lever arm of asking about the very, very distant future that makes even small uncertainties pan out to absolutely ginormous distances in the distant future.

Dwarkesh Patel
I think earlier you said, "I wouldn't be that crazy." But also, "It's not as easy as a Dyson sphere." What are we talking about here? How much energy would it take to...

Adam Brown
The energy requirements are probably pretty small, much more than we can currently make in our particle colliders, but much smaller just in terms of MC squared than the energy in your body. For example, the energy is not going to be the hard bit. The hard bit is going to be concentrating it together in a really small little bubble that's shaped exactly right in order that it doesn't form a black hole, expands in just the way that you want it to expand, and lands in the vacuum that you're aiming for.

So it's more going to be a control issue than just a pure energy issue.

Dwarkesh Patel
But you think this is just table stakes for distant descendants who are colonizing the stars?

Adam Brown
It's not inconsistent with the known laws of physics, which means that it's just engineering.

Dwarkesh Patel
I feel like that the most sort of a churchy phrase physics can occur is "your proposition is not inconsistent with the known laws of physics."

Adam Brown
Not this.

Dwarkesh Patel
If we lived in a world of intelligent design, and these were the laws we found ourselves with, at a high level, what is the creator trying to maximize? Other than maybe us existing, does there seem like something that is being optimized for? What's going on here?

Adam Brown
If you just throw a dart in laws of physics space, in some sense, you would not. There are some properties of our universe that would be somewhat surprising, including the fact that our life seems to be incredibly hospitable for complexity and interestingness and the possibility of intelligent life, which is an interesting fact. Everything is just tuned just so that chemistry is possible.

Perhaps in most places you would throw the dart in possibility space, chemistry would be impossible. The universe as we look around us is incredibly rich. There's structure at the scale of viruses all the way to structure at the scale of galaxies.

There's interesting structure at all levels. This is a very interesting fact. Now, some people think that actually interesting structure is a very generic property, and if we threw a dart somewhere in possibility space, there would be interesting structure no matter where it hit.

Maybe it wouldn't look like ours, but there'd be some different structure. But really, if you look at the laws of physics, it does seem like they're very well attuned for life. So in your scenario where there's an intelligent creator, then they would probably be — you'd have to say they'd optimized for that.

It's also the case that you can imagine explanations for why it's so well tuned for life that don't involve an intelligent creator.

Dwarkesh Patel
Is there any explanation other than the anthropic principle for why we find ourselves in such a universe?

Adam Brown
Well, you suggested one with an intelligent creator, but the usual one that people like to talk about is the anthropic principle.

Dwarkesh Patel
So is it 99% that basically the reason we find ourselves in a universe like this is the anthropic principle?

Adam Brown
What probability do you put high?

Dwarkesh Patel
Well, what probability do you put on, anthropic principle is key to explaining why we find ourselves in the kind of universe we find ourselves in?

Adam Brown
I think it's going to depend on what quantity you're asking me about. So if you ask me, you know, 99% of the matter in the solar system lives in the sun or on Jupiter, and yet we live seems to live in this really weird corner of the solar system, why is that? I'm pretty confident that the answer to that is anthropic.

If we lived in the center of the sun, we'd be dead. And so one should expect intelligent life to live in this weird place in parameter space. So that's perhaps my most confident answer to that question.

Why do we live where we live? Then if we start talking about different constants of nature, we start getting different answers to that question. Why is the universe tuned such that the proton is just a tiny bit more stable than the neutron?

That seems like that's begging for an anthropic answer. Of course, if that's true, that demands that there be different places somewhere in the multiverse where in fact the neutron is slightly heavier than the protons, decay to neutrons rather than vice versa, and people just don't live there. So if you want to go down that road, you end up being naturally drawn to the existence of these variables scanning over space.

Dwarkesh Patel
Is there some way for the anthropic principle to exist that doesn't involve these bubble universes?

Adam Brown
Yes, all you need is that there are different places in some larger possibility space where these quantities scan, where they take different values. Bubble universes are just one way

to do that. We could just be different experiments, simulations in some meta-universe somewhere.

Dwarkesh Patel
What part of this is the least logically inevitable? Some theories seem to have this feeling of like, "It had to be this way." And then some are just like, "Why are there these 16 fields and hundreds of particles?" What part of our understanding of physics?

Adam Brown
I would say that there are three categories. There are things like quantum mechanics and general relativity that are not logically inevitable but do seem to be attractors in some sense. Then there are things like: the standard model has 20 fields, and it has a mass of the neutrino. Why do those masses of the neutrino have the values that they have? The standard model was just fine before we discovered that the neutrinos have mass in the 1990s. And those just seem to be totally out of nowhere. A famous Nobel Prize-winning physicist said about the muon, in fact, longer ago than that: "Who ordered that?" They just seem to be there but without any particular reason.

And then there are these quantities that are somewhere in the middle, that are not logically necessary but do seem to be necessary for life as we know it to exist.

Dwarkesh Patel
How confident are we that these different properties of different universes would actually be inconsistent with intelligent life?

Adam Brown
That's a great question. This line of thought starts to be a skeptical response to the anthropic principle. An example that sometimes people use is a puddle that's sitting in some depression in the ground reflects on how wonderful the universe is, that this depression in the ground seemed to have been made the perfect shape for the puddle to exist. And our view would have said, "No, the reason the puddle has that shape is because it isn't self-adapted to the hole in the ground."

Maybe no matter what the laws of physics, there would be something that emerged there. Certainly, if you go to the bottom of the sea, or in nuclear reactors, or in various other places, this kind of life will find a way. Philosophy seems to be adapted at least there, where it's very different from the surface of the Earth where we find ourselves. And yet they're able to be, certain life is able to live in undersea vents and is able to adapt itself to those environments.

I think I basically buy that life is quite adaptable, but whether life is adaptable enough that a universe with a cosmological constant that ripped it apart every microsecond, that seems implausible to me. Or even closer to home, the center of the sun. It's not clear exactly whether we can get intelligent life living at the center of the sun, even though that has the same laws of physics as us. It just has a different environmental variable.

Dwarkesh Patel
What is the most underappreciated discovery in cosmology in our lifetime?

Adam Brown
In the 2000s and before, we very carefully studied the cosmic microwave background. This, what's sometimes called the echo of the Big Bang, and the inhomogeneity to it, the fact that it's not quite the same in every direction.

Doing that discovered a super interesting fact that was definitely not known in my lifetime, which is the quantum origin of all of the structure we see in the universe. So if you look out in the universe, the density is not the same everywhere. The density on Earth is much more than in interplanetary space, which is itself much more than in intergalactic space, and the center of the sun is all the more denser. It is inhomogeneous; it is not the same.

If you look back to the early universe, it was considerably more homogeneous. It was homogeneous to 1 part in 10^5 or 10^6. Almost everywhere, every point had almost exactly the same density.

So then there's an easy part and a hard part. The easy part is understanding how if you have very small inhomogeneities, how they grow into large inhomogeneities. That's already quite well understood by classical physics. Basically, the idea is this: if you have a place that's denser and a place that's less dense, then the gravitational force pulls stuff towards the high-density stuff. So if you have a small inhomogeneity, they naturally grow under that effect where they just gravitationally fall towards the denser thing. If you start seeded with small inhomogeneities, that will grow large inhomogeneities, and that's well understood.

The thing that we now understand much better than we did is where those small inhomogeneities come from. Why, just after the Big Bang, was the universe not perfectly homogeneous? Because if it was perfectly homogeneous, there's no opportunity for anything to grow.

We now understand with a high degree of confidence something that we didn't understand, which is that those inhomogeneities were seeded by quantum fluctuations. When the universe, just after the Big Bang, was considerably smaller than it is today, the effects of quantum mechanics were correspondingly more important. Those quantum fluctuations produced tiny little fluctuations in the density of matter in the universe. And all of those tiny little one-part-in-a-million fluctuations grew into all of the structures you see in the universe: all the galaxies, you, me, everything else.

Dwarkesh Patel
Is it a meaningful question to ask what level of structure each individual discrepancy corresponds to, each individual 1 in 10^5 part? Is it a galactic supercluster? Is it a galaxy?

Adam Brown
It depends. We believe that these were generated during a period we call inflation, very poorly understood, very early in the universe. And there were fluctuations made not just at one scale in those days, but at all scales, or many, many scales.

There were fluctuations made at a scale that nowadays corresponds to 10% of the distance across the visible universe, all the way down to structures that were inhomogeneities that were much, much smaller scale that correspond to a galaxy today, all the way down to—now, this is speculation, but in some models of inflation, there were tiny inhomogeneities, very small-scale inhomogeneities that would give rise to primordial black holes, like tiny little black holes left over from the Big Bang. There's no actual evidence in terms of observational evidence, no strong observational evidence for those, but those are a possibility. That's allowed by our theory, and people think about them and look for them.

Dwarkesh Patel
What makes General Relativity so beautiful?

Adam Brown
I think general relativity is really an extraordinary story. It's pretty unusual in the history of physics that you, to first approximation, just have one guy who sits down and thinks really, really hard with lots of thought experiments about jumping up and down in elevators and beetles moving on the surface of planets and all the rest of it, and at the end of that time writes down a theory that completely reconceptualizes nature's most familiar force and also speaks not just to that, but speaks to the origin and fate of the universe and almost immediately achieves decisive experimental confirmation in the orbits of astronomical observations or the orbits of planets and the deflections of lights during eclipses and stuff like that. It's a pretty beautiful theory, and it completely changed our idea of gravity from being a force to just being an artifact of the curvature of spacetime, actually.

Dwarkesh Patel
So this is actually a good point to chat about your actual day job. So there's these open debates about the kind of reasoning that these LLMs do. Does it correspond to true reasoning, or is it something more procedural? And it sometimes gets into a definition game.

But this is maybe a good way to test our intuitions here. The kind of thing that Einstein was doing, where you start off with some thought experiments, you start off with some seeming conceptual inconsistencies in existing models, and you trace them through to some beautiful unified theory at the end, and you make incredibly productive use of these intuition pumps, that kind of reasoning. How far are our AIs from that?

Adam Brown

I have heard it said, and I kind of agree with this, that maybe the very last thing that these systems will be able to do, these LLMs will be able to do, is, given the laws of physics as we understood them at the turn of the last century, invent general relativity from that. So I think that's probably the terminal step. And then once it can do that, if it can do that, then there won't be much else to do as far as humans are concerned.

It's pretty extraordinary, I mean, particularly coming from a physics background in which progress is pretty slow, to come to the AI field and see progress being so extraordinarily rapid day by day, week by week, year by year. Looking at it, it certainly looks like these LLMs and these AI systems in some sense are just interpolators, but the level of abstraction at which they're interpolating keeps going up and up and up and we keep sort of riding up that chain of abstractions. And then, presumably, from a sufficiently elevated point of view, the invention of general relativity from Newtonian physics is just interpolation at some sufficiently grandiose level of abstraction.

That perhaps tells us something about the nature of intelligence, human intelligence, as well as about these large language models. If you ask me how many years until we can do that, that is not totally clear, but in some sense, general relativity was the greatest leap that humanity ever made. And once we can do that, perhaps in 10 years, then we will have fully encompassed human intelligence.

Will it be of the same character as what Einstein did? Clearly, there are many disanalogies between human intelligence in these large language models, but I think at the right level of abstraction, it may be the same.

Dwarkesh Patel

Do you see early examples of the kind of thing it was? Obviously not at that level of difficulty, but you just start off with, hey, here's something funny, go think about it for a while. Is there something especially impressive you see when you kind of run that kind of experiment at the moment?

Adam Brown

These systems tend to be doing more elementary material than that. They tend to be doing undergraduate-level material. Yes. I haven't seen anything that jumps out to me like inventing generative relativity or even a toy version of that.

But there is, in some sense, creativity or interpolation required to answer any of these problems. Where you start with some science problem, you need to recognize that it's analogous to some other thing that you know, and then sort of combine them and then make a mathematical problem out of it and solve that problem.

Dwarkesh Patel
Do you think AI mathematicians, AI physicists will have advantages over humans just because they can by default think in terms of weird dimensions and manifolds in a way that doesn't natively come to humans?

Adam Brown
Ah, you know, I think maybe we need to back up to in what sense the humans do or don't think natively in higher dimensions. Obviously, it's not our natural space. There was a technology that was invented to think about these things, which was, you know, notation, tensor notation, various other things that allows you to much using just even writing, as Einstein did 100 years ago, allows you to sort of naturally move between dimensions. And then you're thinking more about manipulating these mathematical objects than you are about thinking in higher dimensions.

I don't think there's any sense, I mean, in which large language models naturally think in higher dimensions more than humans do. You could say, well, this large language models have billions of parameters. That's like a billion-dimensional space. But you could say the same about the human brain, that it has all of these billions of parameters and is therefore billion-dimensional.

Whether that fact translates into thinking in billions of spatial dimensions, I don't really see that in the human. And I don't think that applies to an LLM either.

Dwarkesh Patel
Yeah, I guess you could imagine that if you were just seeing like a million different problems that rely on doing this weird tensor math, then in the same way that maybe even a human gets trained up through that to build better intuitions, the same thing would happen with AI. It just sees more problems. It can develop better representations of these kinds of weird geometries or something.

Adam Brown
I think that's certainly true, that it is definitely seeing more examples than any of us will ever see in our life. And it is perhaps going to build more sophisticated representations than we have. Often in the history of physics, a breakthrough is just how you think about it, what representation you do.

It is sometimes jokingly said that Einstein's greatest contribution to physics was a certain notation he invented called the Einstein summation convention, which allowed you to more easily express and think about these things in a more compact way that strips away some of the other things. Penrose, one of his great contributions, was just inventing a new notation for thinking about some of these space times and how they work that made certain other things clear.

So clearly coming up with the right representation has been an incredibly powerful tool in the history of physics and many incredibly large developments, somewhat analogous to coming up with a new experimental technique in some of the more applied scientific domains. And one would hope that as these large language models get better, they come up with better representations, at least better representations for them that may not be the same as a good representation for us.

Dwarkesh Patel
We'll be getting somewhere when you ask Gemini a question and it says, "Ah, good question. In order to better think about this, let me come up with this new notation."
So we've been talking about what AI physicists could do. What could physicists with AI do? That is to say, are your physicist colleagues now starting to use LLMs? Are you yourself using LLMs to help you with your physics research? What are they especially good at? What are they especially bad at?

Adam Brown
What physicists don't do, or don't productively do, is just say, "LLM, please quantize gravity for me. Go." That doesn't get you anywhere. But physicists are starting to use them in a big way, just not for that.

More of an assistant rather than agent. Three years ago, they were totally useless. No value whatsoever in them. Low-hanging fruit uses include doing literature search. So if you just say, "I have this idea, what are some relevant papers?" They're great at that, and semantically greater than any other kind of search.

The other thing that they're extremely useful for now that they were useless for before is just as a tutor. There is a huge amount of physics that a physicist would be expected to know that has already been done. No human has ever read the whole literature or understands everything, or maybe there isn't even something that you feel you should understand, or you once understood that you don't understand.

I think the very best thing in the world for that would be to phone up a colleague, if you knew exactly who to phone, they'd probably be able to answer your question the best. But certainly, if you just ask a large language model, you get great answers, probably better than all but the very best person you could phone. They know about a huge amount, they're non-judgmental, they will not only tell you what the right answer is, but debug your understanding on the wrong answer.

So I think a lot of physics professors are using them just as personal tutors. And it fills a hole because there are personal... If you want to know how to do something basic, it's typically very well documented. If you want to know quite advanced topics, there are not often good resources for them.

Talking to these language models will often help you debug and understand your understanding. It'll explain to you not only what the right answer is, but what you thought was wrong. I think it'll be a pretty big deal, sort of analogous to the way that chess players today are much better even when they're playing across the board without the benefit of a computer, just having been able to be tutored by chess machines off the board. This is the same: you want to understand this thing about group theory, go and ask the machine and it'll explain it to you and it won't judge you while it's doing it.

Dwarkesh Patel
So there's an interesting question here. Clearly, these models know a lot, and that's evidenced by the fact that even professional physicists can ask and learn about fields that they're less familiar with. But doesn't this raise the question of... We think these things are smart and getting smarter.

If a human that is reasonably smart had memorized basically every single field, and knew about the open problems, knew about the open problems in other fields and how they might connect to this field, knew about potential discrepancies and connections, what you might expect them to be able to do is not like Einstein-level conceptual leaps, but there are a lot of things where just like, "Hey, magnesium correlates with this kind of phenomenon of the brain. This kind of phenomenon correlates with headaches. Therefore, maybe magnesium supplements cure headaches." These kinds of basic connections...

Does this suggest that LLMs are, as far as intelligence goes, even weaker than we might expect given the fact that, given their overwhelming advantages in terms of knowledge, they're not able to already translate that into new discoveries?

Adam Brown
Yes, they definitely have different strengths and weaknesses than humans. And obviously one of their strengths is that they have read way more than any human will ever read in their entire life.

I think maybe again the analogy with chess programs is a good one here. They will often consider way more possible positions, Monte Carlo research, than any human chess player ever would. And yet, even at human-level strength, if you fix human-level strength, they're still doing way more search. So their ability to evaluate is maybe not quite as natural as a human.

The same I think would be true of physics. If you had a human who had read as much and retained as much as they had, you might expect them to be even stronger.

Dwarkesh Patel
Do you remember what the last physics query that you asked an LLM was?

Adam Brown
Well, a recent one was I asked it to explain to me the use of squeezed light at LIGO, which is a topic that I always felt like I should understand, and then tried to explain it to somebody else and realized that I didn't understand it. And went and asked the LLM.

That blew me away, that it was able to exactly explain to me why what I was thinking was incorrect. So, why do we use this particular form of quantum light in interferometers used to discover gravitational waves?

The reason that's a good topic is perhaps because it's an advanced topic, not many people know that, but it's not a super advanced topic. There are, out of a physics literature of millions of papers, there have got to be at least a thousand on that topic. If there was just a handful of papers on a topic, it's typically not that strong at it.

Dwarkesh Patel
Do you reckon that among those thousand papers is one that explains why the initial understanding or thought you had about it was wrong? Because if it just intuited that, that is actually quite like, that's pretty fucking cool.

Adam Brown
I don't know the answer. That is an interesting question. I think it might be able to debug even without that. If you do much simpler things like give these language models code, it will successfully debug your code even though presumably no one has made that exact bug in your code before. This is at a higher level of abstraction than that, but it wouldn't surprise me if it's able to debug what you say in that way.

Dwarkesh Patel
It does falsify a lot of stories about them just being fuzzy search or whatever. Scott Aaronson recently posted about the fact that GPT-4 got a B or an A- on his intro to quantum computing class, which is definitely a higher grade than I got. And so I'm already below the waterline.

But you teach a bunch of subjects, including general relativity at Stanford. I assume you've been querying these models with questions from these exams. How has their performance changed over time?

Adam Brown
Yeah, I take an exam I gave years ago in my graduate general relativity class at Stanford and give it to these models, and it's pretty extraordinary. Three years ago, zero. A year ago, they were doing pretty well, maybe a weak student, but in the distribution. And now they essentially ace the test.

In fact, I'm retiring that. That's just my own little private eval. It's not published anywhere, but I just give them this thing to follow along how they're doing, and it's pretty strong. They, you know, maybe it's easy by the standards of graduate courses, but it's a graduate course in general relativity, and they get pretty much everything right on the final exam. That's just in the last couple of months that these have been doing that.

Dwarkesh Patel
What is required to ace a test? Obviously, they probably have read about all the generality textbooks, but I assume to ace a test, you need something beyond that. Is there some way you'd characterize physics?

Adam Brown
Physics problems, compared to math problems, tend to have two components. One is to sort of take this word question and turn it, using your physics knowledge, into a math question and then solve the math question. That tends to be the typical structure of these problems. So you need to be able to do both.

The bit that maybe only LLMs can do, and wouldn't be so easy for other things, is step one of that: turning it into a math problem. I think if you ask them hard research problems, you certainly can come up with problems that they can't solve. That's for sure.

But it's pretty noticeable, as we have tried to develop evaluations for these models, that as recently as a couple of years ago, certainly three years ago, you could just scrape from the internet any number of problems that are standard, totally standard high school math problems that they couldn't do. And now we need to hire PhDs in whatever field, and, you know, they come up with one great problem a day or something. The difficulty, as these LLMs have got stronger, the difficulty of evaluating their performance has increased.

Dwarkesh Patel
How much do they generalize from these difficult problems to not only that domain of physics but just generally becoming a better reasoner overall? If they just see a super hard GR problem, are they better at coding now?

Adam Brown
Generally, you see positive transfer between domains. So if you make them better at one thing, they become better at another thing across all domains.

It is possible to make a model that is really, really, really good at one very particular thing that you care about. And then at some stage, there is some Pareto frontier and you start degrading performance on other metrics. But generally speaking, there's positive transfer between abilities across all domains.

Dwarkesh Patel
We've got these literally exabytes of data that we collected from satellites, telescopes, and other kinds of astronomical observations. Typically in AI, when you have lots of data and you have lots of compute, something something large model, great discoveries. Is there any hope of using these exabytes of astronomical data to do something cool?

Adam Brown
Yeah, great question. People are trying that. There's an effort, Shirley Ho and Flatiron, which is basically that exact plan.

They take the pipeline of all of the data that comes out of these astronomical observatories, they plug them into a transformer, and see what happens. You can come up with all sorts of reasons in advance why that might not be something that will work, but you could also come up with reasons in advance why large language models wouldn't work, and they do. So I'm very curious to see what happens.

The dream there would be that there are lots of things hidden in the data that no human would ever be able to tease out. And that by doing this, you could just revolutionize the amount of... These astronomical observatories are incredibly expensive. If we can just have a computer better parse all of the data from them in a way that no human ever could, that would be a tremendous improvement.

These things are very good at finding patterns, and maybe they'll find patterns that are not particularly interesting to a human.

Dwarkesh Patel
Okay, so going on the GR thread again, maybe one advantage these models have is, obviously, you can run a lot of them in parallel, and they don't get fatigued or dazed. And you could imagine, again, naively, you would imagine some sort of setup. I assume you're doing much more sophisticated things, but naively, you could imagine a setup where, look, it seems like special relativity, which is something that maybe is easy to understand, is just like you start off with, let's just randomly select a couple of observations. Obviously, they were randomly selected, but you know, and let's just think about what's going on here for a while. Let's just do a bunch of chain of thought for a year or so.

And you could just imagine doing this and doing some sort of best of n across a thousand different randomly selected parts of the current model of the universe and just seeing at the end of it which one comes up with some especially productive line of thought.

Adam Brown
Yeah, I think that could be productive. One challenge in that would be, how do you evaluate whether you had a good theory at the end? That's going to be the tricky bit.

For things that are most easily parallelized are things in which, if you get the right answer, it's clear you got the right answer. You know, perhaps things in NP, one might say. Whereas in this case, is special relativity... How would your computer know when, if it generated special relativity, that it was onto a winner?

There are various ways in which it could know. It could check that it was mathematically self-consistent and various other facts. But the evaluation is going to be a tricky part of this pipeline that you might wish to set up.

Dwarkesh Patel
Is there no experimental way that you could detect time dilation or something?

Adam Brown
There is an experimental way that you could detect time dilation, but that would involve sending out probes or doing something in the real world. Whereas I thought you were just trying to run this in a data center.

Dwarkesh Patel
But now, today, we have these exabytes of information, so you could just have some sort of ability to search or query, like, I've come up with this theory.

Adam Brown
I think maybe this is a philosophical difference, where you maybe think that the way that a theory is good is that it best matches the data with some loss minimization. That's not always how new theories, particularly revolutionary theories, come up.

There's this famous fact: even when they were moving from a geocentric worldview to a heliocentric worldview, it was so beautiful, the theory, by the time they were finished with the epicycles. I mean, not beautiful. It was so ornate by the time where these planets were moving around the sun but moving on epicycles, that actually the data didn't any better fit the heliocentric worldview than the geocentric worldview, especially since they didn't properly understand the ellipticity of the Earth's orbit around the sun.

So it wasn't. Why does one theory replace another? One reason is obviously that it's more consistent with the data, but that's by no means the only theory.

And if you just optimize for being consistent with the data, you're going to end up with—if you optimize only for being consistent with the data—you're going to end up with epicycles. You're not going to end up with some beautiful new conceptual thing. Part of the reason people like these new theories is that even though they're maybe not better at matching the data, they are more beautiful, and we'd have to teach—and that's been a reliable guide in the history of science—and we'd have to teach these LLMs beauty.

Dwarkesh Patel
This actually raises an interesting question, which is, in some sense, we have the same problem with human scientists, right? And there's all these people who claim to have a new theory of everything. I guess there's not an easy verifier that everybody agrees to, because some people call them cranks, other people think they're geniuses. But somehow we've solved this problem, right?

Adam Brown
Well, we've sort of solved it. We haven't solved it in the same way that, if you have some new sort algorithm that you claim is faster than everybody else's, a sort algorithm doesn't need to be any dispute about that. You can just run it and see.

Physics is not the same way. It is definitely the case that there's a number of people who think they have great theories, and there are even perfectly respectable people who are professors at prestigious universities who have very different opinions about what is and isn't a worthwhile direction to be exploring. Eventually, you hope that this gets grounded in experiment and various other things.

But the distance between starting the research program and the community reaching consensus based on data and other considerations can be a long time. So, yeah, we definitely don't have a good verifier in physics.

Dwarkesh Patel
Even if we did someday get superhuman intelligence that could try to find all the remaining sort of high-level conceptual breakthroughs, how much more room is there for that? Basically, was it just like 50 years of "here's all the really advanced-grade physics," and now we just bog through additions to the Standard Model?

If you look at Nobel Prizes, year after year, they get less and less—at least in physics, they tend to get less and less significant. And in fact, this year, the Nobel Prize in physics was awarded to Hopfield and Hinton for their work in AI. So apparently, maybe a taste of things to come.

Adam Brown
I don't think there's a reason — I don't think we should be pessimistic about that. I think there could easily be room for completely new conceptualizations that change things. I don't think it's just turning the crank going forward.

I think new ways to think about things have always been extremely powerful. Sometimes they're fundamental breakthroughs; sometimes they are breakthroughs in which you even take regular physics. This is a story to do with renormalization that maybe is a little too technical to get into, but there was a sort of amazing understanding in the 1970s about the nature of theories that had been around for forever—or for years at that stage—that allowed us to sort of better understand and conceptualize them.

So I think there's good reason to think that there's still room for new ideas and completely new ways of understanding the universe.

Dwarkesh Patel
Do you have some hot take about why the current physics community hasn't—I mean, cosmology is maybe a very notable exception, where it does seem like the expected value of the light cone is switching back and forth.

Adam Brown
Well, if you take particle physics, I think it's because we were a victim of our own success, is that we wrote down theories in the 1970s, and those theories were—it's called the Standard Model—and those theories were too good, in the sense that we won.

In the sense that we could predict everything that would come out of a particle accelerator, and every particle accelerator that's ever been built, and every particle accelerator that's likely to be built given our current budgetary constraints. So particle physics—I mean, there were some questions around the edges, but this model that we wrote down in the '70s and into the '80s basically completely cleaned up that field.

We wish to build bigger, more powerful particle accelerators to find stuff that goes beyond that, but basically, we won, and that makes it difficult to immediately—if you get too good, then it's hard to know where to push from there. That's as far as particle physics is concerned.

Dwarkesh Patel
Is there some—so it sounds like the problem with these colliders is that the expected entropy is not that high of like—the reason it's not that useful is because we kind of have some sense of what we'd get on the other side. Is there some experimental apparatus that we should build where we, in fact, do have great uncertainty about what would happen, and so we would learn a lot by what the result ends up being?

Adam Brown
Well, the problem with particle colliders is, in some sense, that they got too expensive. And CERN is tens of billions of dollars—a small number of tens of billions of dollars—to run this thing.

Dwarkesh Patel
You could build AGI with that money, right?

Adam Brown
It's super interesting how everybody talks about how academics can't possibly compete with the big labs, but the cost of CERN is larger than the cost of big model training runs by a lot. So that's just academics pooling their money. That's an interesting fact.

They got so expensive that it's difficult to persuade people to buy a new one for us that's even bigger. It's a very natural thing to do, to build an atom smasher that just smashes things together to higher energy. It's a very natural thing to see what comes out.

People were perhaps somewhat disappointed with the output of the LHC, where it made the Higgs, which was great, and we found it, but we also expected it to be there. And it didn't make anything else, any of these more fanciful scenarios or anything basically unexpected. People had speculated we'd see supersymmetry there, or we'd see extra dimensions, and basically that was a null result.

We didn't see anything like that. I would say we should definitely build another one if it was cheap to do so, and we should build another one once AGI has made us all so rich that it's cheap to do so. But it's not the obvious place to spend $50 billion if you had $50 billion to spend on science.

Often it's these smaller experiments that can look for things in unexpected places. A decade ago, there was BICEP, which is a reasonably cheap, tens of millions of dollars, experiment at the South Pole that thought it had seen some hints in the cosmic microwave background of gravitational waves. That would have been revolutionary if true.

Not worth doing BICEP if it cost $10 billion. Definitely worth doing BICEP if it costs $10 million. So there's all sorts of experiments like that, often observational.

Dwarkesh Patel
What is the value of seeing these primordial gravitational waves?

Adam Brown
Oh, it gives you hints. You're just examining the night sky very closely and seeing hints of what happened at the Big Bang. This is a sort of different approach to doing high energy physics.

Why do you want to build a big collider? You want to build a big collider because the bigger the collider, the more high energy you can smash things together with. And Heisenberg's uncertainty principle says that high energy means short resolution.

You can see things on very small scales. That's great, except the cost to build them, there's some scaling laws, and the scaling laws are not particularly friendly.

There is another sort of approach that one might say, which is, you know, there was a ginormous explosion that happened, which was the Big Bang. If you imagine, if we look out in the universe, it's expanding. If you sort of play the tape backwards, it's contracting.

Eventually it all contracts at 13.8 billion years ago in the Big Bang. And so that's a very big particle collider indeed. And so by just examining very closely the Big Bang and its aftermath, we're able to hopefully probe some of these quantities that are very difficult to probe with particle colliders.

The disadvantage is that you can't keep running it and adjust the parameters as you see fit. It's just like one thing that happened once, and now we're having to peer backwards with our telescopes to see what happened. But it can give us hints about things that would be inaccessible with any future collider.

Dwarkesh Patel
Is there any information about the distant past that is, in principle, inaccessible?

Adam Brown
Probably not in principle. Something happened to the universe in its evolution, which is that the very early universe, just after the Big Bang, was opaque to light. We can only see light past about 300,000 years after the putative Big Bang.

Before that, everything's so dense. It's like just a dense plasma that light just gets absorbed by. It's like trying to look through the sun.

And so we cannot see directly anything from before 300,000 years. Nevertheless, we can infer lots of stuff that happened from before 300,000 years. In fact, looking at that light, what's called the cosmic microwave background that was emitted at that time, we infer lots of stuff about just due to the patterns of anisotropies that we see in the sky.

We can infer a great deal about what was happening earlier. Most of our confidence about modern cosmology comes from a number of experiments that, starting in the '80s but accelerating in the 2000s, really very carefully measured that anisotropy and allowed us to infer stuff before that. At the information theoretic level, there's nothing inaccessible.

Dwarkesh Patel
I guess that makes sense. Conservation of information. Maybe you will tell me that that also isn't true.

Adam Brown
Well, that's a great question. I mean, there's been a lot of debate in the black hole context about whether information is conserved by black holes, but the modern consensus is that it is.

Dwarkesh Patel
All right, Adam, what are your tips for hitchhiking?

Adam Brown
Oh, good question. So I hitchhiked a bunch around America and Europe. I've done Oxford to Morocco, when I moved from Princeton out to Stanford, I hitchhiked a bunch of other times, down to New Orleans, various other places.

I think probably the biggest tip for hitchhiking is to stand in a good place. Some counterparty modeling. Imagine the person who's picking you up.

They need time to see you, to evaluate you and to decide they're going to pick you up and then to safely stop. And that all needs to happen. So stand somewhere where people can see you, possibly at a stoplight, and where there's a place for them to safely pull over.

Dwarkesh Patel
How do you model the motivations of people who pick you up? What are they getting out of this?

Adam Brown
I think it's different for different people. I think about 20% of people will just always pick up hitchhikers no matter what. Even if I was dressed very differently and presented very different, I think some people would just pick people up no matter what.

I basically fall into that category now. It's hard coded into my brain that I will 100% pick up hitchhikers always under all circumstances. Just because enough people have generously picked me up down the years that I just feel as though it's my duty and sort of not subject to a cost benefit analysis.

Just it's in there. Many other people are evaluating you and just trying to decide what you're in for. Some people are lonely and want somebody to talk to.

Some people have just a spirit of adventure and find it exciting to pick people up. Certainly it's not a representative cross section of people. I would say there's definitely a selection bias in who picks you up.

They tend to be more open and more risk tolerant.

Dwarkesh Patel
And what was your motivation for that? Were you just in need of a car or what was going on?

Adam Brown
No, I enjoy meeting people. I enjoy the experience of meeting people and the weird episodic sense of which, just, you never know what's going to happen. I think I have a very high tolerance for ambiguity, and I enjoy that.

Dwarkesh Patel
What was the percentage of, "We just had a normal conversation, they went in the general direction I was going, and that was that," versus, "I've got a crazy story to tell about X incident"? What percentage is each?

Adam Brown
I think some people are just totally normal people. Families moving their child to college, and you get there and you help them move some stuff into the dorm room just to, just to thank you, all the way through to absolutely wild cases. Probably 20% are just like, this is one of the craziest things that ever happened in one way or another.

Dwarkesh Patel
Any particular examples of the wildest things?

Adam Brown
Oh, yeah, huge. I mean, it's just absolutely a fire hose of wild things happening. I could tell so many stories.

I remember once there was a trucker who picked me up in the desert outside Salt Lake City and who drove me to Battle Station, Nevada. And who, as we were talking—the truckers are always, in fact, the most interesting of all. It's typically illegal or anyway in violation of their employment contract for them to pick people up, so those guys are really, and it's always guys, are really pushing the envelope in terms of picking you up. The truckers often will say, "You're the first person I've had in my cab in 20 years of trucking or something," and then they tell you about 20 years' worth of things that have been on their mind. So, I'd say that those are often the really interesting ones.

As I said, there was this one in Utah who just talked from the moment I got into the cab until we got to Nevada. I kind of got the feeling that he had sort of excess mental capacity and that this was his, you know, he was now just gonna dump it on me. And he was telling me all about his life, and I remember this very well, how his brother-in-law thought he was a loser, his sister's husband. But like, now he had the hot fiance, so who was the loser?

And then just sort of gradually over the course of the six hours, it just suddenly occurred to me that his fiance was doing advanced fee fraud on him. The whole thing was some ginormous, and he was being scammed by his fiance. And very unfortunately for them, they tried to execute the scam while he had me in the cab, and he never had anyone in his cab.

So now he had me in his cab, and they were trying to do some fraud on him. And I was able to, they had some wheat factory in Wales, United Kingdom, that they had some British High Court document saying that he was entitled to if he paid off the lien on it. There was some long, complicated story that was totally flagrantly false.

I kind of felt like I had a moral obligation to him to break the news to him. On the other hand, we were in the middle of nowhere in Nevada, and it was clearly a very important part of his personality that this was so. So I kind of waited until we got close and said, is it possible that your fiance is being scammed by these people? You know, sort of raised the notion of scamming.

And he was willing to intellectually entertain the possibility. And then we got a bit closer. Is it possible that you were yourself being scammed by your fiance? And then he was like, "No, no, no, it can't be." And he had all these documents to show that it was all legit, and they were just sort of, to somebody from a British legal background, transparent forgeries.

He did eventually accept it and was just crying on my shoulder in some truck stop. It was quite a high pathos moment. And then said, uh, this happened before.

And it turned out he'd previously been scammed in the same way or a similar way through somebody he'd met through the same match.com profile. That was his lucky profile because, you know, people kept messaging him through it. So we, you know, we talked through that and worked through that and, like, I felt in some ways I'd been his guardian angel and, uh, but he, you know, he'd also been my guardian angel and picked me up in the middle of the desert, so there was some, there was some great exchange there.

Dwarkesh Patel
That's crazy. I hope he closed down that profile.

Adam Brown
I hope so. I mean, we, you know, I did chat to him about that possibility and he wasn't fully bought in on it, but, uh, yeah.

Dwarkesh Patel
What's the longest you've been stranded somewhere?

Adam Brown
That would probably be one time in Richmond, Virginia, in some not particularly good neighborhood, trying to hitch out of there. I think that was about a day, which is really bad. That's really bad.

Sometimes, if you get a good spot, that's worth a thousand miles. Just don't, don't give it up just for a short hop anywhere. If you get a bad spot, get out of there on any means necessary because there's, because there's probably a thousand X variance in how high quality hitchhiking spots are, I would say.

Dwarkesh Patel
How did you find the time to like get stranded for a day at an end?

Adam Brown
In terms of intensity, it doesn't really take that much wall clock time, as we say. Coast to coast is, um, you know, like a week or so. It's pretty fast because you don't, you're not yourself driving, in that sense, it's easier.

You do have to wait, and, you know, there is definitely high variance how long you can be. But in terms of sort of incidents per minute, it's, it's a pretty good way to see the world. And you see such a cross section of people who I might never, never otherwise meet, and such a sort of high variance cross section.

Everything from sort of idle millionaires cruising around the country looking for adventure to people who just got out of prison to, in one memorable incident, well, it eventually transpired as we were going along that they were, uh, they had, they were actually just teenagers and I didn't, somehow didn't clock that when getting in the car. And they, they had stolen the family car and were, were driving west, um, without a plan. And that, yeah, there I gave him a talk, I was talking to, and, uh, bought them dinner and some, some life advice. So that was some fun stuff I got.

Dwarkesh Patel
Did you make them call their parents?

Adam Brown
I did make them call their parents, yes. Or, you know, heavily encouraged them to call their parents.

Dwarkesh Patel
Is there a luck to get the professor?

Adam Brown
Yeah, none of these people typically realize that, uh, you know, your academic background never really comes up in conversation, typically. I mean, sometimes it does, but typically that's, that's not the nature of the conversations.

Dwarkesh Patel
Was there any time you felt particularly unsafe?

Adam Brown
I have definitely felt more unsafe picking up hitchhikers than I have hitchhiking. Maybe I just got lucky, but picking up hitchhikers, there it tends to be, um, you know, no one really picks up hitchhikers, uh, anymore, and there's definitely a selection effect on who's hitchhiking.

Dwarkesh Patel
Right.

Adam Brown
I've definitely felt more in risk of my life with hitchhikers I picked up than I ever did hitchhiking. But, you know, it's possible I just got lucky. You don't see the other branches of the wave function.

Dwarkesh Patel
What are the other interesting insights from just getting this random cross section?

Adam Brown
Yeah, all sorts of facts. A lot of people just like to talk. There's a lot of, a lot of people out there, and I like to talk too, so it's mutually beneficial.

Dwarkesh Patel
Well, the truckers, I imagine are especially so.

Adam Brown
Those guys are interesting. They're all cheating their logs. They have certain logs about how long they can travel for, and at least every single one who's ever picked me up has all been,

in some way or another, gaming the system of their logs about how long they're allowed to drive for and playing games with time zones.

They're smart people, and they just have a lot to say and don't really have anybody to say it to. So they're very grateful.

Dwarkesh Patel
What are they especially insightful about?

Adam Brown
They tend to have listened to a huge number of audiobooks. They have a ginormous amount of information stored in their brain, but nobody to tell it to. Also, many of them tend to have had unlucky romances at some stage in their past that they've never really gotten over or spoken to.

I really feel as though many of them would do well to speak to a therapist. But you are the therapist in that case. In many ways, people will tell you things.

Frequently people will say things like, "I've never told anybody else this in my life before." That's common, not just the truckers, other people as well. Sometimes it's families picking you up, and so they're not going to say that.

Often it's just single people picking you up, and they'll say, "I've never said this before to anyone else in my life." And they'll tell you some story of their life. I do think it's an exchange, and they're also getting quite a lot out of the conversation.

I remember one case going to New Orleans. Somebody just meant to only take us, I think it was just some state trooper had come along in South Carolina and was going to arrest us because it's illegal in some states to hitchhike in North Carolina. And so I was like, "Okay, just take the next ride."

And it was just 10 miles down the road. He ended up getting so into it that we ended up driving maybe 1,000 miles out of his way by the time we'd gone. He'd had this, we were having great conversations, just absolutely wonderful time, and he just wanted to keep going and going and drive us through the night.

Then we ended up going through the Deep South in the middle of the night and arriving near New Orleans around dawn. He'd had a father who had been in the military, but he'd kind of had a difficult relationship with. He ended up going and visiting his father's grave in Baton Rouge, never having done that in the 20 years since his father died.

But just as this sort of turned, I mean, he just was driving along expecting to go home, and then it just turned into this sort of spiritual quest for him. Stuff like that can be pretty gratifying. It's also sort of cheating.

You're not, in my way of thinking about it, meant to be taking people out of their way. They're meant to be going where they're going, and you go with them, and they take you no further. But in this case, I think he needed to go there, so that was good for him.

Dwarkesh Patel
Did you stay in contact with any of the people you hitchhiked with?

Adam Brown
Typically, no. I would almost consider it poor form to do so. But actually, there was one lady who came to stay in New York later.

She was going down to Haiti to sort of be a doctor there. She was a doctor, and so I stayed in contact with her a bit. But typically it's just the nature of the interaction is that you have this sort of beautiful moment in time together, and then that's it.

Dwarkesh Patel
Any other tips that somebody should know? I mean, should they do this anymore, given that it's largely uncommon and so uncommon types of people might pick you up?

Adam Brown
I think it used to be very common in the United States. It's still reasonably common in Europe. It used to be very common in the United States, and then there were some mass murderers who drove the popularity down by targeting hitchhikers.

Maybe this is just pure cope. In my mind, you need to worry about that less because if you are a mass murderer, it's really not a high expected value strategy to cruise around looking for hitchhikers since there are so few of them. But that just might be pure cope in my head.

I've never refused a ride for safety grounds, but I would, I hope I would if necessary. Sometimes you would refuse a ride because somebody is only going a short distance, and you're in a good hitchhiking spot. It's kind of bad karma to refuse a ride, but sometimes you should do that.

Other tips: don't write your exact destination on your sign. Write the sort of direction in which you're going. The reason is maybe twofold.

One, a lot of people, if they're heading towards that place but not going to that place, will not stop because they think, "Oh, I'm not going to wherever it is, I better not. I'm not going there, so I won't pick you up," even though you'd very much appreciate a partial ride there. The other reason is if you do want to decline a ride, it's certainly a lot easier to do so.

Dwarkesh Patel
If the person says, "Oh, I'm going to that city."

Adam Brown
Right, that's hard. Where if they say they go into that city and you've written something more vague on your sign, then it's maybe easier to decline a ride. If you want to get out of the car, the classic, if you get in and you feel unsafe, is to say that you're carsick because even serial killers don't want vomit in their car.

So that's a good reason to get out, and then you just say, "Okay, I'll just stay here." That's another trick. I've never had to deploy that.

Dwarkesh Patel
Oh, I was just about to ask.

Adam Brown
No, I've never had to deploy that. Typically, it's pretty, there's a moment of anxiety in the first minute. But then after a minute, it's clear that everybody is, they're also anxious about you. In many ways, you can tell that they're quite nervous about you. And then after a minute, it's clear that everybody's, if not a sensible human being, then at least a safe human being. And everything's super relaxed for the rest of the ride, typically.

Dwarkesh Patel
Any other strange people who picked you up that come to mind? Not necessarily strange, but just memorable.

Adam Brown
So many different kinds of people. I remember there was one seemingly very successful cowboy, driving some fancy truck in Wyoming. He had a big herd of cattle and all the rest of it, and was asking me what I do.

At that time, I was doing cosmology, so I tried to explain to him. It had no connection with anything. He just didn't understand a word I was saying all the way through.

Eventually, we landed on the fact that the stars in the sky are just like the sun, only much further away. This was a fact that, in his life up to that stage, he had just never encountered.

That was extremely gratifying because he was blown away by that fact. He wasn't intellectually incapable of understanding it. He just never, in his 50 years of existence up to that moment, ever heard that fact.

His mind was just totally racing. This was reorienting his picture of his place in the universe. "The universe must be so big if there are stars out there!"

He phoned his wife, who I think was somewhat less excited, and then took me to a gun store and bought me lunch. He was a rancher, seemingly a very successful rancher based on everything about him. He had some prize, high-quality bulls that were some rare kind of high-quality bulls. I can't exactly remember the details, but he just never really contemplated what the night sky meant for him.

Dwarkesh Patel
There's a Sherlock Holmes story where Holmes learns that actually the sun is the center of the solar system.

Adam Brown
Oh, interesting.

Dwarkesh Patel
Watson tells him this, and Holmes is like, "Why did you tell me this? I try to reserve mental space for things that are actually relevant to my work. Now I have to forget this."

Adam Brown
A Hitchhiker's Guide to the Galaxy.

Dwarkesh Patel
What did you learn from studying the first-hand accounts of the Nagasaki bombers?

Adam Brown
During the pandemic, my landlord had a big library, and I just started reading some books in the library during deep lockdown. There was some sort of enigmatic statement in some book about the history of Japan.

Dwarkesh Patel
Where do you stay that your landlord has a library?

Adam Brown
I live in a house that used to belong to the chair of the English department at Stanford, and then it was inherited by his grandson who rents it to me. He has a very extensive library.

I was going through it during the first lockdown and came across this super enigmatic statement in a book about the history of Japan. I was super fascinated by it and started, for reasons that I'll explain in a moment, then just became obsessed for a few months on reading absolutely everything I could about the bombing of Nagasaki.

It's the most recent nuclear weapon ever to be set off during wartime. It was reasonably controversial because people questioned whether we should have done it or not. That wasn't the question I was looking at. The question I was looking at wasn't, "Should they have ordered it to be done?" but, "Were the people who did it even following orders?"

It's a pretty wild story that I certainly didn't know before any of this happened. It was never meant to be a mission to Nagasaki. It was meant to be a mission to bomb Kokura, a different Japanese city, but they got there, and it was clouded over.

They had very strict instructions: "Do not bomb unless you can see the target." That was the order.

They got to this other city, passed over a bunch of times, and they couldn't see the target because it was covered in clouds. Then they went to their secondary target, Nagasaki, and it was again covered in clouds, and they did a whole bunch of passes.

They'd made various mess-ups beforehand, including getting lost. They'd made a number of personal flying mistakes on their part that meant that they didn't have enough fuel once they got to Nagasaki to carry the bomb back to base. They probably would have ended up in the ocean had they tried.

They were extremely motivated at the time. This was the only nuclear weapon that existed in the world. We'd had two, and then it went down to one. Now there was one, and they were just about to drop it in the ocean and lose it.

According to the official account, after having done all this, on the third and final pass over Nagasaki, there was a miraculous hole in the cloud that suddenly opened up, and then they dropped it. That story is a bit suspect, if for no other reason than that they actually missed. Little-known fact, they missed Nagasaki.

They were aiming for one point, and they hit another point that was on the other side of the hill, such that the original thing they were aiming for was reasonably untouched by comparison, considering the fact that a nuclear weapon had been dropped. They missed by much more than you would miss if you were doing visual bombing, and they had been told to do visual bombing.

There's this kind of suspicion that they were doing a little bit of radar bombing against direct orders. So is it possible that 50% of all of the nuclear weapons ever dropped in combat were, in fact, dropped against direct orders? If true, that's a pretty striking fact about nuclear war, since people are somewhat worried with nuclear war that someone will launch nuclear weapons without being ordered to do so.

It does kind of look like 50% of all the nuclear weapons ever dropped in combat were dropped against direct orders. When they got back, Curtis LeMay was going to court-martial them and was super mad, but then the war ended, and they didn't want to do it for PR reasons.

I just ordered and found every account ever written by every person. It was super fascinating to do that because all these different people had completely non-overlapping lives. Some of them were on the Manhattan Project and were there as observers and later won Nobel Prizes for physics. Some of them were just people who were just there for one moment.

Dwarkesh Patel
So Louis Alvarez was on the plane?

Adam Brown
There was typically a physicist, a representative of the Manhattan Project, on the plane just in case. So Louis Alvarez was someone there. He actually wasn't on the Nagasaki mission. He was on the Hiroshima mission.

But in his biography, he's like, "They said they saw a hole in the clouds. I don't think I believed them." So that was, I think, one of the hints. It was maybe reading his autobiography at some stage, that was one of the big hints.

The other people insist there was. But what's super clear is that, whether or not there was a hole in the clouds, and probably there was a hole in the clouds, just because of some of the technical things to do with their discussion, though it's definitely not obvious.

What's clear is that whether or not there was a hole in the clouds, they certainly had decided in the cockpit on that final run that, no matter what, they were going to drop it. So even if there wasn't a hole in the clouds, they had decided to drop the nuclear weapons against direct orders.

Dwarkesh Patel
And had they written, basically, "Oh, we totally saw a hole in the clouds, but even if we hadn't, we would have dropped it."

Adam Brown
That basically is. Yeah. So different people write different things.

Dwarkesh Patel
How did you end up on the plane?

Adam Brown
There are about ten people on these planes. Did any of them say? Not all of them were, you know, some of them were some ways away from where the action is happening. There's the bombardier who says that he saw a hole in the clouds. There's the pilot who says something.

But everyone has their own different perspective, and some of the perspectives are just totally, this is something that I guess I'd always been told by my history teachers but never really appreciated until I'd done this 360-degree view of history, that people can describe the same events and just, they have flatly inconsistent memories of each other.

Nobody who was on the plane said that they faked the hole in the cloud story, but some people who were on the plane said they were determined to drop the bomb no matter what, and they were highly incentivized to do this, because if had they not done it, they'd have probably, as it was, they only barely made it back to their emergency landing spot in Okinawa. They would have definitely ended up in the drink, and certainly the bomb would have ended up in the drink had they not done it.

So I don't know. I'm not a professional historian, and maybe there'll be a difference of opinions, but it's clear there was something highly sus about at least 50% of all the nuclear weapons dropped in combat.

Dwarkesh Patel
The interesting thing is that the reason nuclear war was averted in other cases is also because they refused to follow direct orders, right? So in this case, or in the case of Petrov, he didn't report the seeming sighting of Nuke Storm America, and that obviously contradicts orders.

Adam Brown
Yeah, there's nuclear insubordination in both directions.

Dwarkesh Patel
That's right.

Adam Brown
There's the good kind, where they maybe should drop the bomb according to their orders and refuse to, and then there's the other kind.

Dwarkesh Patel
I also want to ask, so you've had not only one remarkable career but two remarkable careers. In physics, you're a close collaborator of people like Leonard Susskind, and you've done all this interesting research. Now you're helping do the reasoning work that Google DeepMind's working on in AI. Is there some chronology you have in your head about how your career has transpired?

Adam Brown
Oh, I don't impose narratives on it like that. It's certainly a very big contrast between doing physics and writing retail papers, as it were.

Dwarkesh Patel
Retail.

Adam Brown
Doing one-by-one writing physics papers and then doing AI, which moves just tremendously faster, and trying to contribute to wholesale production of knowledge in that way.

They have very different impacts in terms of counterfactual impact. In physics, you write some papers, and you're like, had I not written that paper, no one would have written that paper for years or ever, perhaps. Computer science doesn't feel like that. It feels like if you didn't do it, someone else would do it pretty soon thereafter. On the other hand, the impact, even a few days of impact in computer science, these things are going to change the world, hopefully for the better, to such a large degree that that's much bigger than potentially all the physics papers you ever wrote.

Dwarkesh Patel
That's interesting you say that about, you feel that physicists are not fungible in the same way. The story about why physics has slowed down is usually that, in fact, there isn't any low-hanging fruit. The idea that you would discover something that somebody wouldn't have written about for many years to come. I had a couple of double negatives there, but basically, we found all the things that you can just write a paper about. You're not just going to think about something and find something that somebody else wouldn't have written about otherwise.

But here, you're saying the field that's moving way faster, which is computer science, that's the one where all these people are going to come up with your algorithms if you hadn't come up with them yourself. And it's physics where if you had more Leonard Susskinds and Adam Browns, you would have much faster progress, potentially.

Adam Brown
Well, partly there's just so many more people working on the problems in computer science than there are in physics. Just the number of people is part of what makes the counterfactual impact.

Dwarkesh Patel
How many theoretical physicists are there versus how many people are working on AI research?

Adam Brown
AI research around the world? I don't know how many people are in research, but it's like thousands and thousands and thousands. The amount of matter is 100, 200, 300.

Dwarkesh Patel
Really?

Adam Brown
Well, in the narrow domain of high-energy theoretical physics. There are many more physicists than that if you include people more generally, but they're sufficiently specialized. I mean, that's partly part of the reason is that it's a much more specialized field. So in a very specialized field, the number of people who would actually write that paper is a much smaller number.

Dwarkesh Patel
How much do you ascribe the slowness of physics to these kinds of things that are just intrinsic to any field that is as specialized and as mature versus to any particular dysfunctions of physics as a field?

Adam Brown
Yeah, we look back on the golden era of physics from the 1900 through 1970s or something as a period when things happened. I do think there is a low-hanging fruit aspect to it. We already talked about how the Standard Model is so successful in terms of particle colliders that it's just hard to make rapid progress thereafter. So I don't really see it as a dysfunction of the field so much as being a victim of our own success.

Having said that, does physics have fads? Does physics have fashions? Does physics have any of these other things? Absolutely, it does. But quite how much counterfactual progress we'd make if that weren't true, I don't know.

Dwarkesh Patel
How well-calibrated are the best physicists?

Adam Brown
It doesn't necessarily pay to be well-calibrated, and that incentive structure is perhaps reflected in the poor calibration of many of the best physicists. First of all, because physics is a sufficiently mature field, all the good ideas that look like good ideas have already been had, or many of them. Where we're at now is the good ideas that look like bad ideas.

So in order to motivate yourself to get over the hump, get over the barrier, and actually explore them, you need a little bit of irrational optimism to ride out the initial discouraging things that you'll discover as you go along.

I would say that typically theoretical physicists are not particularly well-calibrated and tend to be in love with all their own theories and make highly confident predictions about their own theories. Before the LHC turned on, there were certainly a lot of high-energy theorists making extremely confident predictions about what we'd see at the LHC, and it was typically their own favorite particle that we'd see. While I'd love to have found supersymmetry, it would have, in some sense, felt somewhat unjust to reward the hubris of people making overconfident and poorly calibrated predictions. So, yeah, that's definitely a thing that happens.

Dwarkesh Patel
But I wonder if poor calibration on the individual level is somehow optimal on the collective level.

Adam Brown
I think that's basically right. The same is kind of true in other domains of life as well. Of course, with startups, if you were properly calibrated about how likely your startup is to succeed, maybe you wouldn't do it.

But it's good for the ecosystem that certain people are willing to give it a go. I think it's good for the ecosystem and perhaps bad for the individual to be well-calibrated.

Dwarkesh Patel
Another topic I know you studied a lot is how one might mine a black hole.

Adam Brown
Oh yeah, right. I read a paper about that. Very good.

Dwarkesh Patel
Tell me about it.

Adam Brown
Okay, so what do we mean by "mine a black hole?" Mining a black hole means taking energy out of a black hole that used to be in a black hole. Obviously, if our distant descendants have used up all of the energy in stars and everything else, the black hole might be the last thing they turn their eye to.

Can you get energy out of black holes at all? The old story, pre-1970s, is no. A black hole is one way: matter falls in, it never comes out, it's stuck.

The thing that Hawking and Bekenstein discovered in the 70s is that once quantum mechanics is involved, that's not true anymore. Once quantum mechanics is involved, in fact, energy, even without you doing anything, starts to leave black holes. The problem, as far as our distant descendants will be concerned, is that it leaves black holes extremely slowly.

So if you took a solar mass black hole, same mass as the sun, just collapsed to form a black hole, there'll be this little quantum, what's called Hawking radiation nowadays, little quantum Hawking radiation in which the energy will leach out again very, very slowly. The temperature of a solar mass black hole is measured in nanokelvins, a very low temperature. So the energy leeches out when something that cold, so cold you couldn't even see it in the cosmic microwave background, it leeches out incredibly slowly back into the universe.

And that's bad news because it means the energy comes out super duper slowly. So the mining question is, can you speed that up? A solar mass black hole, if you don't help it, will take about 10 to the 55 times the current age of the universe to have given out all its energy back into the universe.

Can you make that faster? There were these proposals stretching back a few decades that you could do what's called mining black holes, where we see the Hawking radiation that escapes when we are a very long way away from the black hole. But actually, mathematically, it's known that much of the Hawking radiation doesn't escape.

It just sort of makes it a little bit out of the black hole and then falls back in again. And there was this proposal that you could kind of reach in with a mechanical claw, obviously not crossing the horizon, because otherwise, you've lost the claw and you're somewhat counterproductive, but just outside the horizon, just grab some of that Hawking radiation and just drag it a long way away from the black hole and then feast on it or do whatever it is you want to do with it. In that way, you could mine a black hole.

You could speed up the evaporation of a black hole by a huge factor. So in fact, the lifetime would no longer go like the mass cubed, like it does with just unaided Hawking radiation, but would scale like just the mass, so considerably faster for a large black hole.

So this was these proposals and what I had a somewhat pessimistic contribution to the story, which is that the existing proposals did not work. They didn't work to speed it up. And in fact, you can't speed it up.

You can't get down that M cubed down to M. You can't, in fact, get it anything less than M cubed. It still scales like the mass cubed. The length of time you need to wait to get all the energy out of a black hole still scales like the mass cubed.

And what goes wrong is ultimately a material science problem. So this scoop that comes down really close to the horizon, now, from one point of view, that's just like a space elevator, albeit a very high-performance space elevator. Space elevators, you'll remember, are these ideas for how we might get things off the surface of the Earth without using rockets.

The idea is that you have some massive orbiting object sort of very long way away, beyond geostationary orbit, and then you dangle off that a rope down to the surface of the Earth, and then you can essentially just climb up the rope to get out. That's the space elevator idea. And already around Earth, it's hitting pretty hard material science constraints.

So if you want to make a space elevator, the trouble with making a space elevator isn't so much supporting the payload that you're trying to have climb up. It is merely just the rope supporting its own weight because each bit of the rope needs to support not only its own weight but also the weight of all of the rope beneath it. So the tension that you require keeps getting more and more and more as you go up.

At the bottom, there is no tension effect. It doesn't even touch the Earth. It's not like a compression structure that's like a skyscraper that's pushed up from below. It's a tension structure that's held up from above.

But as you go up, because you need more and more tension, you also need to make the rope thicker and thicker and thicker. And if you try and on Earth or around Earth, build a space elevator out of steel, say, it just doesn't work. Steel is not strong enough.

You need to keep doubling the thickness until, by the time you get to geostationary orbit, the thickness of the steel rope is more than the size of the Earth. Like, the whole thing just doesn't work at all. But carbon nanotubes are this material that we discovered that are much stronger than steel.

So, in fact, around Earth, carbon nanotubes will just about work. If we can make them long enough and pure enough, then they will be strong enough that we will be able to build a space elevator around Earth in, you know, maybe sometime in the next century, that you only need a couple of doublings of the thickness of the carbon nanotubes along its entire length. So carbon nanotubes work great around Earth, but they are totally inadequate for black holes.

For black holes, the critical material science property you need for this rope is the tensile strength to mass per unit length ratio. It needs to be strong, high tensile strength, but low weight, light, low mass per unit length. And that's the critical ratio.

And carbon nanotubes is 10 to the minus 12 or something on that scale. And that is simply not strong enough at all. In fact, what I showed in my paper is that you need a tensile strength to weight ratio that is as strong as is consistent with the laws of nature.

So, in fact, the laws of nature bound this quantity. The finiteness of the speed of light means you cannot have an arbitrarily strong rope with a given mass per unit length. There is a bound set by the C squared in some units that bounds the maximum possible tensile strength that any rope can have.

Any rope, in fact, that has that, or an example of a rope that has that, is a string. So a string is, I mean, a fundamental string from string theory is an example of a hypothetical rope that is just strong enough to saturate that bound, that strength bound. And then the problem is the following.

The problem is that if you have a rope that saturates the bound as strong as any rope can be, it is just strong enough to support all of its own weight exactly on the edge there, with exactly no strength left over to support any payload it might wish to carry. And that's ultimately what dooms these mining black holes, you know, these rapid mining black hole proposals.

Dwarkesh Patel
And what happens if you try to make the rope stronger?

Adam Brown
Well, you can't. One example of a thing that goes wrong is the speed of sound in a rope goes up with the tension and down with the mass per unit length. And if you try and use a rope that's stronger than this or some hypothetical rope, you would find that the speed of sound is greater than the speed of light. And that's a pretty good indication.

Dwarkesh Patel
What is the speed of sound?

Adam Brown
So, if you just take a rope stretched between you and me and ping it, there will be little vibrations that head over towards you. Those vibrations are subluminal. If it's just a normal

rope, they move at the speed of light. For a string or something that saturates null energy condition, it would be faster than the speed of light.

That would be an example of why there's something wrong with that proposal.

Dwarkesh Patel
So, it just happens to be the case that the rope cannot mine black holes. I think we've mentioned a couple of other bounds like this where there's no principled reason you might have anticipated ex ante why there would be such a bound that prevents something that just gets in our way, but it just so happens to be this way.

Does this suggest that there's some sort of deeper conservation principle we'd be violating? And then the universe conspires to create these engineering difficulties which limit that?

Adam Brown
Yes, nothing is ever a coincidence. Usually, from the perspective of the story I just told to do with mining black holes, it's not clear what exactly will be broken about the universe if you could mine black holes somewhat faster than we can. There are other ways of thinking about it in which, if you could make a string that was strong enough to actually do it, if you could make a rope that was stronger than this bound, various other things would go wrong.

There are various symmetry arguments that that can't happen. But often, it turns out, if we have these bounds, that there's something that sort of saturates the bound or gets very close to the bound. And that's a sign that you're on the right lines with some of these bounds.

Dwarkesh Patel
On the right lines in what sense?

Adam Brown
As in, if you have a bound but you can't think how to get close to the bound, that's usually an indication that you need to think closer. Because often these bounds, if you're clever enough, there's a way to get to the bound.

There's no rule that it has to be so, but that's often the case. Someone will come up with a bound, and there will be a gap between the bound and how close we can get. Usually, more ingenuity will take you up to the bound.

Dwarkesh Patel
I guess the thing I'm curious about is why it would be the case that such a bound would exist in the first place. And how often do you run into these things? Basically, are you expecting

to discover something in the future about why it had to be this way, that you can't mine black holes?

Something would be violated that tells us something important about black holes, that they can't be mined, and it's deeper than the tensile strength of the string that would be required to mine it.

Adam Brown
Yeah, good question. I started these investigations because it offended my intuition for various information theoretic reasons, the idea that black holes could be mined with parametric speed ups. When I thought harder about it, the reasons why I thought that couldn't happen didn't really make sense.

So in this particular case, maybe someone will come up with a reason. I don't actually have a particularly strong reason why they can't be mined anymore, except that they can't.

Dwarkesh Patel
Okay, so we can't get the material out of the black hole at a pace that would make it reasonably useful to us. What can we do with black holes? What are they good for?

Adam Brown
If you have a small black hole, you can get stuff out of them more rapidly. The temperature of a black hole is inversely proportional to its size. So one thing that people have talked about with black holes is using them to extract all of the energy from matter.

As you know, most chemical reactions are pretty inefficient. You burn gasoline and you extract, as a function of the rest mass of the gasoline that you started with, one part in 10 billion of energy from the gasoline that you started with. So that's bad from the point of view, you know, you have MC squared worth in a gallon of gasoline. You've got a full MC squared worth of energy in there, and you can only get out one part in 10 to the 10. That's a pretty unsatisfactory situation.

Roughly speaking, the reason that all chemical processes are so inefficient is that they only address the electromagnetic energy in the electrons. A very small fraction of the electromagnetic energy in an electron in atoms is stored in the electromagnetic interaction between the electrons and between the nucleus and the electrons. Most of it is stored in the nucleus itself, in the strong nuclear forces, and particularly in the rest mass of the protons and neutrons that constitute it.

So you can do much better if, instead of doing electromagnetic interactions, you use nuclear interactions that can probe the energy in turning protons into neutrons. That's why nuclear power plants are so much more efficient on a per mass basis than chemical power plants like coal plants or gas plants because you're getting a much higher fraction. Best case scenario, you're getting one part in 10 to the three or 10 to the four of the rest mass of the uranium that you start with, you're extracting as energy.

But even there, even in that process, it's still only absolute best one part in a thousand of the rest mass. And the reason is that you are using where much more of the energy is stored, which is the strong and weak interactions between the protons and the neutrons. So much more is available to you.

But still, at the end of whatever the process you finish with there, there's a number that will be conserved, and that is what's called the baryon number. So it's the total number of protons plus the total number of neutrons. You can transmute protons into neutrons or vice versa in nuclear processes, which is part of the reason they're so much more energy than things that just affect the chemistry.

But still, most of the energy is stored in the rest mass of the protons and the neutrons. And you want to get that, and nuclear processes conserve that. Beta decay will maybe turn a proton into a neutron or vice versa, but the total number of protons plus neutrons is not changing. And so therefore 99.9% of the energy is inaccessible to you.

So what you need to do to get that energy and try and get most of the MC squared out of the matter that you have, what you need to do is use a process that eats baryon number, in which you can start off with a proton and a neutron and end up with no proton or neutron. Instead, all of that energy is unleashed in high energy radiation that you can use for your own purposes.

So electromagnetic interactions won't do that. Strong interactions also won't do that. Weak interactions won't do that. The only force of nature that will do that, with a small caveat, the only force of nature that we know that will do that is the gravitational interaction.

And so it is a property of black holes that you can stand outside the black hole and throw protons and neutrons into the black holes, and then it'll process it and then spit out photons at the end in Hawking radiation and gravitons, which is going to be slightly annoying to have to capture, and neutrinos. But they're there in principle, and in principle, you could capture them. So one thing that black holes might be technologically useful for in the future is you start off with a much smaller black hole than what I've described, than the size of the sun.

Dwarkesh Patel
Be very careful about making sure it doesn't grow.

Adam Brown
You can be super duper careful and throw in protons and neutrons and then get out photons. In principle, if you could capture everything that's emitted from the black hole, including the gravitons and the neutrinos, that gets rid of the baryon number conservation problem. It allows you to build power plants that approach 100% efficiency.

And by 100%, I mean, not the way we measure gas turbine efficiency, where we talk about the total available chemical energy in the gas. I mean, 100% of the MC squared of the entire gas you're putting in.

Dwarkesh Patel
Although, if you consider our cosmic endowment, we're not exactly lacking for mass.

Adam Brown
We have a lot of mass. On the other hand, we also have plans for our future that involve exponential growth, and eventually we will run low on that mass. Not that many doublings before using up the whole galaxy, so you want to use it carefully.

Dwarkesh Patel
Let's talk about black holes. How much information can a black hole store?

Adam Brown
That's a great question. That has been a very productive line of thought. The answer to that question goes back to Hawking and Penrose.

You could even ask another question, which is: how much information can anything store?

Dwarkesh Patel
Can we back up? Why do we ask this question of black holes in particular? How often do we ask, "How much information can the Sun store?" Why are we interested in how much information a black hole can store?

Adam Brown
Well, it turns out that that's been an incredibly productive line of thought. And it also turns out that that is the main fact that we're most confident about about quantum gravity.

So the two great theories of 20th century physics: gravity, Einstein's theory of the curvature of spacetime and gravity; and quantum mechanics, the theory of the very small, to do with Heisenberg's uncertainty principles and atomic spectra. Gravity tends to make itself felt at the very large scale, and quantum mechanics tends to make itself seen at the very small scale.

These are the two most beautiful theories of 20th century physics, the two things that we should be most proud about that we discovered in the early 20th century. It was noticed pretty early on that these two theories seem to be inconsistent with each other. The most obvious ways to try and reconcile quantum mechanics and gravity break. You can't really shove them together.

And this is a problem if you think that the world should be comprehensible, that there should be some theory that is consistent, that describes the world. So this has been a big project in theoretical physics over the last few decades, trying to understand how we can take Einstein's general relativity and quantum mechanics and make them meld together in a mathematically and physically consistent manner.

It's tricky, in part because there's very little experimental guidance because general relativity tends to make itself felt at large scales, quantum mechanics at small scales. So trying to find a place where they meet in the middle, and it must be that they do meet, but trying to drag that out with experiment is very tricky.

This has been a big project, trying to figure out how to do this. Einstein spent some years unsuccessfully doing this in the later, less productive part of his career.

This project of trying to unite these is something that a lot of people have thought a lot about. String theory comes out of this project, a number of other lines of thought.

There is, however, one fact about that merger that we are most confident about, and about anything about the merger. And that exactly returns to this question of how much information you can store in a given region of spacetime. And in fact, how much region.

And the answer to that involves black holes. So the answer is how much? If you have a region of a certain area, maybe a sphere of a certain area, and you said, "How much information can you store in that region?" The amount of information you can store, measured in bits, the entropy of that region, is given by the area of that region divided by G, Newton's constant, and H bar, Planck's constant.

So that's how you know that this is something to do with quantum gravity because it involves both G and H bar.

Dwarkesh Patel
Is that the only situation in physics where both of those constants end up being in the same place?

Adam Brown
That is not the only situation. No, anytime you have quantum gravity, they'll tend to be in the same place. And sometimes even when you don't have quantum gravity, but you have the interplay of gravitational forces and quantum degeneracy pressures, those will also end up in those.

But it's in some sense the simplest situation in which it occurs, which is why so much time has been spent thinking about thought experiments to do with black holes. So there was a physicist called Bekenstein who figured out that that should be the answer, the area divided by GH bar. And then Hawking's great contribution to physics was figuring out that it was the area divided by GH bar, but he also got the pre-factor, and the pre-factor was a quarter.

So Hawking figured out that it's a quarter at the area divided by 4GH bar. And this is a super interesting answer. How much information can you store in a given region is given by the area. And in fact, black holes maximize that. Black holes store that amount of information in a given area.

Dwarkesh Patel
But specifically area, meaning surface area?

Adam Brown
Meaning surface area, exactly. The reason that that's such a wild answer, and an answer that's led to all sorts of thought experiments to do with quantum gravity ever since then, is that you might naively think that the amount of information you can store in a region is given not by its surface area, but by its volume.

If I have a hard drive, and I take another hard drive, and another hard drive, and another hard drive, and I keep piling them up, the amount of information I can store on those hard drives scales like the number of those hard drives. That means it scales like the volume of the region in which I'm storing the hard drives. Everything we know about classical thermodynamics tells us that the amount of information should scale like the volume.

Everything we know about non-gravitational physics tends to point in the direction that the amount of information you can store goes like the volume. And yet, this is the most surprising fact that is incredibly generative: once you add gravity to the picture, once you combine quantum mechanics and gravity, the amount of information you can store in a given region, a given sphere, goes like the surface area of that region, not like the volume of that region.

You might think that that can't possibly be right. You might give the following argument: there's some region, and I'm just going to keep adding more and more hard drives to that region. As I make that region bigger and bigger and bigger, the amount of information on those hard drives scales like the number of those hard drives, which goes like the radius of that region cubed.

The thing about the radius of the region cubed is it grows faster at a large radius than the radius of that region squared. So I just told you that the amount of information you can store in a region is given by the surface area, and yet I also gave you a way to make it scale like the volume. So eventually, if I make the region big enough, the amount of information in that volume will be bigger than the bound that I just said.

Therefore, I've ruled out Hawking's, Penrose's, and Bekenstein's bound. What goes wrong with that thought experiment is that eventually, if I make a big enough pile of hard drives, the whole pile of hard drives will undergo gravitational collapse and form a black hole.

Dwarkesh Patel
But then there has to be an experiment—not experimental, but do you have to crunch the numbers then to determine that just before the pile of hard drives would collapse into a black hole, the amount of information stored in that cubic pile of hard drives is less than the amount of information that then gets turned into the surface area of the black hole? Because theoretically, I don't know if I'm getting my mathematicians right, it's theoretically possible that even though the black hole is smaller because it's only the surface area, the cubic ends up being bigger.

Adam Brown
You have to run that calculation. But if you do run the calculation, it turns out that it's nowhere near. It wasn't close.

Dwarkesh Patel
It's not one of those things where they just balance each other out.

Adam Brown
They don't just balance each other out. If I take an online shopping website and I buy a bunch of Western Digital hard drives, and I calculate the information storage capacity of those and compare it to the area of a black hole, I figure out when the pressure in the hard drive would be enough to stop it collapsing to form a black hole. It is nowhere close. It will make a black hole way before it comes close to violating the Bekenstein-Hawking bound.

Dwarkesh Patel
Got it. Okay, sorry. I didn't mean to interrupt. And then you...

Adam Brown
So that's the information storage in black holes. The reason you know that that's also the information storage bound for anything, not just black holes, is that if you had something

that wasn't a black hole that had more information than that in a given region, and you just added matter, eventually that thing itself would collapse to form a black hole. And so it couldn't be the case, just logically, that it had more information than the black hole. It'll tend to...

Dwarkesh Patel
You just hinted at the idea that somehow this is the most productive line of thought that physics has come up with in the last few decades. Why is that? Why does the fact that the area is proportional to the information of a black hole tell us so much about the universe?

Adam Brown
It's been extremely important for our understanding of quantum gravity. It's perhaps the central fact that we know about quantum gravity: the information scales with the area. That fact, which was known since the 70s, was a big hint that became very influential later on.

As understood by Bekenstein and Hawking, it was just a weird fact about black holes, perhaps. But we now understand it as a strong indication of what we call the holographic principle. The holographic principle has been a powerful idea in quantum gravity, and it's the following.

If you took a non-gravitational system in which you ignored gravity, like the pile of hard drives, the information storage would scale like the volume, as we discussed. Whereas in fact, it scales like the area. Another way to say that is if you take a three-dimensional, three-plus-one-dimensional theory in which you have both quantum mechanics and gravity, the information storage scales like R squared rather than R cubed.

That is, it scales as though you had a non-gravitational system in one fewer dimension. So if you had a two-dimensional theory in which there was no gravity, the information stored in a given region would also scale like R squared because the information would be just the two-dimensional volume, as in the area. So in other words, at least as far as information density, the information capacity is concerned, a gravitational theory in three dimensions is like a non-gravitational theory in two dimensions.

Or more generally, a gravitational theory in n dimensions is like a non-gravitational theory in n minus 1 dimensions. So that is a big hint that forms the basis of the holographic principle. It's like gravity eats information. There's less information than you thought there was, than you naively thought there was if you didn't include information.

And so the holographic principle says that maybe that's not just a neat observation. Maybe it is, in fact, the case that for every, or for some quantum gravitational theories, there is another theory that is exactly equivalent to it in one fewer dimension. And so this led to

Maldacena's ADS/CFT correspondence, the gauge gravity duality, which was the most cited paper in high-energy theoretical physics ever, I think, maybe at this stage.

In the late 90s, he wrote down an exact, we believe, an exact duality between a particular theory of quantum gravity, some particular flavor of string theory, and a non-gravitational theory that lives on the boundary of that space.

Dwarkesh Patel
And what problem does it solve if you can model the world in fewer dimensions that doesn't involve gravity?

Adam Brown
This was a very influential paper, and really becomes a tremendous theoretical laboratory for trying to understand the connection between gravity and quantum mechanics. One problem it solves is this: gravity is mysterious, particularly once we improve quantum mechanics in various ways. This is why it's hard to quantize gravity.

But if you can say that this theory that involves both quantum mechanics and gravity is exactly dual—is in some sense the same theory as just an alternative description of a theory in one fewer dimension that doesn't involve gravity—that's great, because we have a much better grasp on how to understand theories that don't have gravity than we do on theories that do have gravity. So, it puts everything on a much clearer footing to have this non-gravitational description, because then you can just use the standard tools of non-gravitational quantum field theory in order to define it and understand it.

Dwarkesh Patel
At one level, I understand that if the information in an area is limited by the information that would be on the surface of a black hole in that region, then you can model the surface area as a two-dimensional object. On the other hand, if I just think about the real world, you're over there and I'm over here, and if I do something here, it's not interacting with you. In order to model that fact, I need to model the dimension in which—the third dimension in which—we're separated.

I guess if I'm actually looking at you through a window pane, maybe I wouldn't have access to that. So, in two dimensions, how do you model that? There's a reason we have the third dimension, right? And how is that modeled if you reduce that dimension?

Adam Brown
Maybe I should just lead with some disappointing news: AdS/CFT was a tremendous conceptual breakthrough in our understanding of quantum gravity and embodied the holographic principle, but at the same time, it doesn't describe our universe. In particular, in AdS/CFT, there is a negative cosmological constant in the gravitational theory, and our

universe, as we discussed before, has a positive cosmological constant. So, it's great because it provides an existence proof of a well-defined theory of quantum gravity—not, alas, in the universe in which we live.

But having said that, it's extremely confusing and was a very impressive result precisely because you might think, how could it possibly be the case that two different theories in two different dimensions could turn out to be equivalent? The answer to your question is, if you have two people who are living in this negatively curved space and talking to each other, what does that look like in this other theory? I say that there's this process going on in the gravitational theory that's dual, which is exactly isomorphic to some process going on in the non-gravitational theory in one fewer dimension.

But what maybe looks very simple in one theory, like you and I chatting back and forth to each other, would look like some complicated plasma physics in the lower-dimensional boundary theory. And so, the complexity of how it looks, which is a better description, does not need to be conserved across the isomorphism. In fact, that's often what we use it for.

We use it to do arbitrage between things that look simple in one theory and things that look simple in the alternative description. We use the fact that things look simple in one to understand the sort of complicated version in the other. In fact, it flows in both directions.

You might naively expect that because gravity is so complicated, we would always be using the non-gravitational theory to understand the gravitational theory. That's not always true. Plasma physics is itself extremely complicated.

There are these big collisions that we do at RHIC in Brookhaven where we smash two gold atoms together and make big fireballs of quark-gluon plasma. It's extremely challenging to calculate what would happen there, and yet people use this duality in the opposite direction to say, even though it looks super complicated with this weird plasma physics in the non-gravitational theory, it actually can simply be understood as some simple black hole property in the gravitational theory.

Dwarkesh Patel
Maybe not AdS/CFT itself, but would some theory which relies on the holographic principle ever be able to account for a world like ours, where, unlike the surface of a black hole, there isn't a boundary because of the positive cosmological constant and it's constantly expanding? Is there some hope that there in fact is a way to have some sort of dual theory to this that somehow describes a boundary?

Adam Brown
People are working on that. That is an open area of research. Ever since the original AdS/CFT was written down, people have been trying to formulate versions of it which have a positive cosmological constant.

It's difficult. Part of the difficulty goes all the way back to Archimedes: it is easiest to formulate a theory if you have a fixed point on which to stand and observe things from a distance. In a universe with a positive cosmological constant, you don't have that.

You don't have that. You're necessarily mixed up with the system because you live in a universe that has only a finite amount of entropy, a finite amount of free energy. There is an inherent limitation to the precision of the experiments you can do.

That just makes things way trickier. So, for that and related reasons, it's a much harder project, but for sure people are working on that.

Dwarkesh Patel
What is the correct conceptual way to think about this? One version is, the boundary is one way to simplify the processes that are actually four-dimensional. Another is—I don't know how we think about this in the context of black holes—but maybe in the context of black holes, no, the information actually is on the horizon. The analogous thing here would be, no, somehow we are on the boundary of the universe somehow. Is there a sense in which one of these interpretations is correct?

Adam Brown
This duality idea, where you have two different descriptions of the same thing, is not new. The AdS/CFT correspondence was not the first such example in physics. It's a common trope in physics that you can have two different descriptions of the same thing, some of which are more useful in one scenario, some of which are more useful in the other scenario, but both are exactly correct.

There are non-gravitational examples in physics that go back a long way. You may then ask, "Which one is right, and which one is not right? Is it actually a CFT that's pretending to have this weird alternative description as a gravitational theory? Or is the gravitational theory correct, and the other one is not correct?"

I think this is more of a philosophical question. My answer would be that if the isomorphism was just an approximation — if it was really one thing and you were just pretending it was the other thing, and that approximation worked in some region of validity and not others — then I would say that one was right and the other one was just an alternative, fanciful description.

That is not our understanding of AdS/CFT as we understand it today. Our understanding is that this is a precise isomorphism. It's not an analogy, it's not a metaphor, it is not an approximation that is valid in some domain and not another.

It really is the case that these two theories are exactly equivalent to each other. And if that's correct, then as a matter of philosophy, I would say those are both equally real. So it's not the case that one is more real than the other; they're perfect simulations of each other.

Are you an AdS dreaming you're a CFT, or a CFT dreaming you're an AdS? I think these are just two completely different, inequivalent descriptions of the same identical physics.

Dwarkesh Patel
Tell me if this is just a question that doesn't make sense. Because when I was, if you try to ask somebody about the quantum many worlds, "Where are the other worlds?" right? And they're just like, "They're in Hilbert space."

"Where is Hilbert space?"

"No, dude, it's just a conceptual thing. Stop asking questions."

Intuitively, it feels like there should be a sense in which there is some physical existence. Either that existence is in this four-dimensional space, or it's in some space that exists on the boundary. Is this just going to lead us into philosophical loops, or is there something that can be said more about it?

And also, in a de Sitter space, in a world like ours, what exactly would the boundary mean?

Adam Brown
There are two components to that question. You have an intuition that if something is real, it needs to be spatially localized, and things that are delocalized in space somehow can't be real. I would say that that's not my intuition.

My intuition is that there can be two completely different descriptions of the same physics, and if it's precise, neither of those is any more real than the other. Things do not need to be spatially localized. You separately asked, what would a version of "where is the boundary theory?" in de Sitter space, since there's no boundary?

That is a great question that people who are trying to generalize AdS/CFT to a universe like ours, that has a positive cosmological constant, wrestle with. There's more than one proposal. Some suggest that the dual theory should live on the cosmic horizon.

So, if you go 5 billion light-years, you can send information to that point and have it returned to you. But on the other hand, there are things that are 100 billion light-years away that we'll never be able to communicate with. There's a boundary between those two — between things that we could, in principle, communicate with and things that we couldn't, in principle, communicate with.

That is the cosmological horizon. Some people who are trying to do a version of holography that works in universes with a positive cosmological constant like to put the second theory there. Other people like to put it in the distant future, in the sort of infinitely distant future.

That's part of the problem: where do we even put that theory? It's not like in our universe where you can just put it spatially infinitely far away and be done with it.

Dwarkesh Patel
If it's spatially finite, then we are currently at the boundary of infinite many other universes that are located, or whose center is located, elsewhere.

Adam Brown
Absolutely. A cosmological horizon is very different from a black hole horizon in this regard. A black hole horizon — there is a point of no return.

If you get closer than that, you fall into the black hole; you're never getting out again. And everybody can agree where that is. For cosmology, there is a point of no return, but the point of no return is relative to a given person.

For each person, there is a different point of no return. And as you say, we live on the boundary just as much as we live on the boundary of — those people live on our cosmological horizon, we may live on theirs.

Dwarkesh Patel
Okay, another philosophical question. There seem to be many theories which imply that there's some sort of infinity or approximate infinity that exists. In quantum many worlds, there are constantly these different branches of the wavefunction spawning off where things are slightly different.

So everything that can possibly happen has happened, including basically the same exact thing. I guess if this bubble universe stuff is correct, it implies a similar picture. Philosophically, should it have some implication on our worldview?

It would be surprising that we learned this much about the universe and then it has no implications whatsoever, right?

Adam Brown
Good question. I think I'm going to say yes and no. I mean, it's clearly, if correct, let's just take the quantum case, which is perhaps even more secure than the cosmological multiverse case.

In the quantum case, it really does look like the default expectation, given everything we understand about quantum mechanics, should be the many-worlds interpretation in which the universe keeps branching off, and there'd be more and more branches. Every time, or almost every time you come to a point of quantum a measurement, we might colloquially say is made that the universe branches, and then every possibility is represented still in the grander wave function.

That's a pretty profound thing to learn about the ontology of the world. If correct, it seems like it should be the default expectation.

And you might say, maybe I don't care about existential risk in our universe because we blow each other up or turn into goo or whatever. Okay, that's sad for us. Maybe our world has vacuum decay, but there are some other branches of the wave function where it's not. And so some other branches will have made different choices in the past, and they're sort of guaranteed somewhere in the branches of the wave function to be a flourishing world. And so I'm not so bothered.

I would say that that's, I'm not going to tell you what utility function you should place on the wave function, but Born is. There's the Born rule in quantum mechanics.

And that tells you that you shouldn't just say, if it's there in one branch, that's just as good as anything else. Born's rule, which is one of the foundational rules in quantum mechanics, tells you how much to care about each branch. You don't care about them equally.

It says that the correct way to calculate the expectation value of anything is to calculate its value in each branch and weight those branches by the square of the amplitude of the weight function, which is some particular quantity, and then add together all of those different answers. So that's a linear answer, which is to say that the total utility of the universe is the sum of the utility in each of these branches appropriately weighted by Born's rule.

So if that's true, you should hope to make our branch as good as possible, just because whatever is going on in the other branch, the total utility is just the sum of what's going on in that branch and what's going on our branch. And so you should try as hard as you can to make our branches as great as possible.

Nevertheless, I do kind of understand that you might have a portfolio theory that seems to be inconsistent with Born's rule, but is somehow intuitive, in which somehow it's not just a linear function on these universes.

Dwarkesh Patel
This would only be applicable if you are a total utilitarian who then there's a sort of very straightforward way in which we can dismiss this and be like, it's one of these. It seems like in physics, there's always these kinds of things where, like, oh, we think we discovered something new, but how would you look at that?

The speed of light is still conserved. And similarly here, like, oh, infinite universes. Ah, but would you look at that? That has no implications on our decisions.

But most people are not total utilitarians. And if you have some very simple thought experiments to illustrate a couple. Suppose that there are two universes and, sorry, two worlds in two different cosmic horizons who will never interact with each other causally, but each one has intelligent life and civilization and beauty and everything we might care about.

If one of the two gets extinguished, I'm like, pretty sad. And this is, both of these make up the entire universe. If both of them get extinguished, I'm more than twice as sad.

There's something to that sort of finality which makes existential risk salient in the first place. And if you agree with that intuition, then I think you should be inclined to think that there is something significant about the fact that in some base reality, like genuinely the story carries forward.

On the other end, if you're somebody who cares about minimizing the downside of people talking about suffering, risk or something. Right. The idea that if it's physically possible to have a universe full of torture, it's actually in fact happening or will happen again is like you could just be like, ah, but the amplitude on that is so small or the square of the amplitude is so small.

Adam Brown
And.

Dwarkesh Patel
The weighted average ends up close to nothing. But I'm like, that really sucks, that's actually happening.

Adam Brown
Yeah, I think there are a number of ways to think about this. I think in part people's intuition is maybe formed in cases like extinction, where if you have an animal that's going extinct, if

half of the animals get wiped out, that's somehow less bad than if both halves of the animals get wiped out.

But that's because they really aren't going to interact in the future. And there's the possibility of the. Those don't have non-overlapping future light cones, the two populations of some possibly extinct animal.

It's also the case that this is a pretty like Born's rule, narrowly defined, does not really have anything to say about this how one should calculate the total utility. It's just more of a sort of the natural utility measure that would come out of this.

Particularly when you get to the cosmological multiverse. I think that these are very difficult questions to answer your intuition that perhaps two different universes in which how we calculate those do we just add together the utility in both or is there some nonlinearity to do with it?

Basically, for the cosmological multiverse, there isn't a particularly good way to decide what the weighting factor should be. We don't have the same equivalent of Born's rule in quantum mechanics. And I think it's at least open for opinions like yours to be to be.

In fact, there should be some better way in which we calculate it. That's not just a linear function of.

Dwarkesh Patel
These different kinds of infinities is there some sense in which some are more fundamental than others. That is, maybe the bubbles are artifacts of what's actually happening on the wave function, or vice versa.

Adam Brown
You're talking about the two kinds of multiverse, the sort of cosmological multiverse and the quantum mechanical multiverse. Yeah, they get very bound up if you try and write down a theory that has both of them.

Because whether there's a bubble there, you're trying to make bubble universes. But what gives rise to bubble universes is often quantum processes. So often you end up in superpositions over there being a bubble universe and there not being a bubble universe there.

And that means that these two kinds of multiverse, the sort of quantum mechanical multiverse and the cosmological multiverse end up getting totally intermeshed with each other.

Dwarkesh Patel
But it sounds like the base reality is still the wave function over all the bubbles in the entire inflaton field or whatever.

Adam Brown
Yeah. So again, we only really properly know how to do quantum gravity and do the counting when there's a negative cosmological constant. As we discussed with ADS cft, in these bubble universes, where there's a positive cosmological constant, it's still somewhat an open question how to do the accounting of what happens and where and how much it should count out.

Dwarkesh Patel
Okay.

Adam Brown
Which is to say we don't know the answer to that question, and your opinion is not ruled out.

Dwarkesh Patel
It's a little bit confusing because in one context, we're laying out very practical—I don't know if you can call black hole batteries practical—but very tangible limitations on the what future, like very distant future descendants, could do with all the matter in the galaxy and so forth. On the other hand, we're like, "Bubble universes as big as our own made somewhere in somebody's lab, maybe." So, basically, how confident are we that the practical limitations we think we know about will actually constrain our future descendants?

Adam Brown
That's a good question. Certainly, some of the possibilities we've discussed so far have different epistemic statuses about how confident we are or are not. And as we also discussed, some of these bounds are somewhat fragile.

Can you communicate faster than the speed of light, for example? Let's just take that as an example bound. We think you can't, according to the laws of science as we understand it. Most physicists would be pretty surprised if it turned out that you could.

Dwarkesh Patel
What is your probability if, like, a million years from now, we are able to communicate faster than light? How surprised are you?

Adam Brown
That is a tricky one. That is a really tricky one. It's only a century that we've thought you can't communicate faster than the speed of light.

A million years is such a radical time that maybe we've sort of dissolved the question into some greater question, and we understand it doesn't even really make sense. I would be pretty surprised. If you make me make a number, I think that there is a greater than 90% chance that in 100 years, we are still limited by the speed of light. There's a 98% chance, if you make me be precise.

Dwarkesh Patel
Okay, so then what are the other constraints on a future civilization that they might care about? If we've got these superhuman intelligences and they're colonizing the galaxy, what are the things they might want to do that they can't do? They probably care about energy. They care about computation.

Adam Brown
Energy limits. We've talked about the efficiency of batteries and extracting energy.

MC squared is the — I'm highly confident that the most energy you can extract from a given piece of matter is MC squared, at least until you start getting cosmology involved. Other limits will be Landauer's limit, or in other words, with a given amount of energy, how useful is a given amount of energy to you? We wouldn't care about having huge amounts of energy if you could get an arbitrary amount of value out of a fixed unit of energy.

We think that that's not true. We think that in particular, if we're going to do computations with it, for example, that there's going to be—and that computation makes errors—that there is a fixed cost of a bit, basically a bit of free energy, in order to correct those errors.

Dwarkesh Patel
And we're confident that there's no way to make computers that don't make errors?

Adam Brown
It is a very interesting question what the fundamental limits on errors are in a computer, how far down they can be pushed. In terms of never making errors, I think that's very unlikely. If for no other reason than there is a minimum background temperature caused by the expansion of our universe—again, it all comes back to the cosmological constant—that gives a very small but non-zero temperature to our universe, that I think will inevitably mean that we make errors.

You might imagine we could just set up some kind of perpetual motion machine that's just thinking happy thoughts over and over again in a quantum computer that never tires and never stops. I think that inevitably the universe would leak in, and there would be errors. But what the minimum error rate is, is not—I don't have a clear answer to that question. Physics doesn't have a clear answer to that question.

Dwarkesh Patel
One question you might have is: What will be the nature of not only the things that our descendants might care about, but what will they be able to produce domestically? What will they want to trade for? If something like alchemy is just super—you know, it's like equals MC squared is all you care about—then it's just like, look, your star system or your galaxy has a certain amount of mass, and you can convert that to energy, and there's fundamentally no reason to trade if there's not that high transaction cost to make it into whatever you want.

On the other hand, if there are some limits, like in fact you had to make galaxy-wide factories, or you had to do these NP-hard calculations that even with a galaxy you can only trace down certain segments of the search space or something, there might be reasons to trade extremely—sort of a pie-in-the-sky question—but how much can we intuit about these kinds of constraints?

Adam Brown
In economics, the theory of comparative advantage only applies if not all resources can be transported. If you can just go in and just disassemble whoever you're doing the comparative advantage with, you might as well just turn them into—apply it all to the party with the absolute advantage. So maybe the same thing would be true in the universe.

I think there are a number of questions in there. For starters, not all energy is equally useful in different places in the universe. If there's a galaxy over there and a galaxy here on this side of the universe, because of the expansion of the universe, if I beamed the energy—if I disassembled that galaxy and tried to send it back here, either by literally sending it on starships or converting it to light and beaming the light back in a laser and then having a big PV here to collect it, or for whatever mechanism—by the time it reached me, there will be a massive redshift.

So keeping it in place is maybe better than just disassembling it and bringing it back home. But there's another question, which is—these are all unknowns to do with both physics and the nature of technology—is the most important thing that all of the value will be created here on Earth, and we just need to get as many resources back here on Earth? And there are super linear returns to scale of having accumulated resources in one place, so we just want to make Earth an absolute paradise? Or do we want to spread—is it in fact sublinear, and we want to spread civilization all the way throughout all of these galaxies?

I think questions like that are going to be important in addressing your question of what the returns to scale are and returns to trade as well.

Dwarkesh Patel
It's like the galaxy in a billion years from now has a certain GDP. What percentage of that GDP do you think is just the end result of computations or confirmation that a computation

has been made? Maybe it's like simulating hedonium that the other side of the galaxy cares about or something.

Adam Brown
Just because it may prove to be so much more efficient to do things in simulation than to do them in the real world, my guess would be a high percentage of that. But maybe that's wrong.

Dwarkesh Patel
If computing is the main thing you hear about, what is going to be the—physically, how will the flops in a galaxy be organized? Will it be as planet-wide computers, as a huge blob the size of a star system? Do we have some sense of —

Adam Brown
This is a super interesting question. It returns to the question we were asking before: with quantum computers, we know, for example, that the amount of quantum computation you can do in terms of the equivalent amount of classical computation when trying to do some factoring algorithm or something grows super-linearly with the number of qubits. In fact, it grows almost exponentially with the number of qubits.

So a 200-qubit quantum computer is much more than twice as good as a 100-qubit quantum computer for certain tasks. But for the tasks that we try and use quantum computers for, that's true. So that line of reasoning might lead you to believe that in the distant future, we will just try, even paying the cost of the redshift and all these other questions, we'll feed all of the energy and free energy back into one central quantum computer.

It will all be about making that central quantum computer as big as we possibly can, even at the cost of inefficiency. On the other hand, there are other kinds of tasks for which actually having a twice-as-big computer is not that much better, or certainly not more than twice as better, than having two smaller computers. In that scenario, it'll be a more distributed setup.

Dwarkesh Patel
I guess in this quantum computer system, you would need to have coherence across this huge, which might not be a practical engineering difficulty for future civilizations, but does seem...

Adam Brown
Either it would need to be co-located, or you'd need to send the quantum coherence out. That's actually not that hard to do. It's a property of photons that they do tend to maintain when they're propagating in the vacuum.

They basically maintain their coherence for a very long way. In fiber optic cables, you reach trouble because they start getting absorbed by the fiber optics after tens of miles. But in the vacuum, you could, in principle, share quantum entanglement across the universe if you did it right.

Dwarkesh Patel
Then wouldn't you expect, when you say a central computer, physically it wouldn't just be like a huge, contiguous...

Adam Brown
Well, it might be because the sort of analog of the classical fact that flops are not the only thing you care about. You also care about bandwidth and interconnects and things like that. So perhaps the same would be true.

I mean, here we're getting into a pretty speculative area, but you could imagine either configuration, either one in which you have a huge number of different quantum computers that are talking to each other via entanglement networks or one in which you just have one big central computer.

Dwarkesh Patel
Final question. Timeline to when you are automated as a physicist.

Adam Brown
Oh, good question. Many of the tasks that I might have performed in the past, I think, are already automated at some level. Until I am totally out of the picture and no longer necessary...

That's probably pretty close to ASI complete. So whatever your timeline for ASI is.

Dwarkesh Patel
Well, I guess the question is, what is yours?

Adam Brown
Yeah, I'm squirming somewhat uncomfortably in answer to that question because I'm not totally sure. I could certainly imagine a scenario in which it's five years.

Dwarkesh Patel
All right. I think that's a great place to close. Adam, thanks so much.

Adam Brown
Thank you. Great to be here.