

Science AMA Series: I’m Anthony Aguirre, Professor of Physics at UC Santa Cruz. I’m searching for other universes, trying to understand black holes, bringing funding to not-quite-crazy physics, worrying

Anthony *Aguirre*¹ *and r/ScienceAMAs*¹

¹Affiliation not available

April 17, 2023

Abstract

I’m a cosmologist specializing in the very, very, very early Universe (cosmological inflation), with interests in black holes and gravity, the foundations of quantum mechanics, the arrow of time, and other fun-and-bewildering things. My favorite subjects are those that sound somewhat crazy, but arise from following our current best physical theories where they lead. The inflationary “multiverse” is a good example, and my recent research mostly concerns observable signatures of other bubble “universes” crashing into ours. In 2006, Max Tegmark and I created the Foundational Questions Institute to support research worldwide on the foundations of physics and cosmology where the questions are so risky, or “big” that conventional funding is hard to come by. If you enjoy thinking about the stuff on this thread, there’s lots of good material at the FQXi community site. Last year, I joined Max and several others in founding the Future of Life Institute, which seeks to minimize the risk/reward ratio in world-changing emerging technologies such as AI and Genetic engineering. Like other universes and black holes, when you talk a lot about these technologies they can sound crazy. But they’re coming! Most recently, with astrophysicist Greg Laughlin, I started the Metaculus project, an effort to create high-quality crowdsourced predictions about scientific and technological impact and breakthroughs, to help us understand what science and technology we should keep our eyes on. (Looking for more users – if you like to predict or think about the future, check it out!) Edit: 5 PST, periodically checking in to answer a few more

[REDDIT](#)

Science AMA Series: I'm Anthony Aguirre, Professor of Physics at UC Santa Cruz. I'm searching for other universes, trying to understand black holes, bringing funding to not-quite-crazy physics, worryin

ANTHONY_AGUIRRE [R/SCIENCE](#)

I'm a cosmologist specializing in the very, very, very early Universe (cosmological inflation), with interests in black holes and gravity, the foundations of quantum mechanics, the arrow of time, and other fun-and-bewildering things. My favorite subjects are those that *sound* somewhat crazy, but arise from following our current best physical theories where they lead. The inflationary "multiverse" is a good example, and my recent research mostly concerns [observable signatures of other bubble "universes" crashing into ours](#). In 2006, Max Tegmark and I created the [Foundational Questions Institute](#) to support research worldwide on the foundations of physics and cosmology where the questions are so risky, or "big" that conventional funding is hard to come by. If you enjoy thinking about the stuff on this thread, there's lots of good material at the [FQXi community site](#). Last year, I joined Max and several others in founding the [Future of Life Institute](#), which seeks to minimize the risk/reward ratio in world-changing emerging technologies such as AI and Genetic engineering. Like other universes and black holes, when you talk a lot about these technologies they can sound crazy. But they're coming! Most recently, with astrophysicist Greg Laughlin, I started the [Metaculus](#) project, an effort to create high-quality crowdsourced predictions about scientific and technological impact and breakthroughs, to help us understand what science and technology we should keep our eyes on. (Looking for more users -- if you like to predict or think about the future, check it out!)

Edit: 5 PST, periodically checking in to answer a few more

[READ REVIEWS](#)

[WRITE A REVIEW](#)

CORRESPONDENCE:

DATE RECEIVED:

March 12, 2016

DOI:

10.15200/winn.145769.98592

ARCHIVED:

March 11, 2016

CITATION:

Anthony_Aguirre , r/Science ,
Science AMA Series: I'm
Anthony Aguirre, Professor of
Physics at UC Santa Cruz. I'm
searching for other universes,
trying to understand black
holes, bringing funding to not-
quite-crazy physics, worryin,
The Winnower
3:e145769.98592 , 2016 , DOI:

Due to the nature of physics of conservation of energy/mass, how likely is it that our universe was "born" from another universe? (thinking of multiverse, where each universe is a membrane) because we must've come from something. I know that there's a " great attractor" which is pulling a lot of galaxies to one mysterious part of the universe but there's also a " great void" where there's not a lot. Is it likely that this is due to influence from other possible universes? Or is it that we still don't quite understand gravity, dark matter or dark energy? Thanks :)

[Nerdork86](#)

Great question! As someone else has noted, energy conservation in general relativity is very subtle. In classical mechanics energy conservation stems from the "laws of physics" (the Lagrangian) being independent of time. But if you have a object moving in a space-time that evolves, then the energy of that object is not conserved. This is one way to understand the cosmological redshift of photons.

There is a way to define the energy of the entire universe, if the universe is closed and finite. But by this definition, the energy is always zero, no matter what you do. So it does not strongly constrain what happens inside the universe. One way of looking at this is that while matter and radiation have positive energy, the gravitational field has negative energy. In a closed universe, these always exactly cancel.

[10.15200/winn.145769.98592](https://doi.org/10.15200/winn.145769.98592)

© et al. This article is distributed under the terms of the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and redistribution in any medium, provided that the original author and source are credited.



So even if the closed universe gets arbitrarily large, with arbitrarily much stuff in it, it can still have zero energy.

More constraining than conservation of energy, I think, is conservation of information, or “unitarity.” If the laws of physics really are unitary, you might imagine that they must describe something at $t < 0$, if $t=0$ is the putative “beginning of the universe”, where classical general relativity breaks down. In other words if you believe there is a theory of quantum gravity that respects unitarity, then the Big Bang singularity is really just a copout and you should think about what happens on the “other side” of that singularity — or you should devise a cosmological model that does not have a big Bang singularity, which is my preferred solution.

How can you convince the average person that subjects such as quantum mechanics/physics etc are real/relevant and not just theoretical thoughts made up by clever people using really big words?

[rollmop1](#)

Physics and quantum mechanics are indeed things made up by clever people using big words. But they also, quite amazingly, describe the actual universe and are incredibly useful. The fact that you can take a question about the real world, translate it via physics into abstract mathematics, find a result, and translate it back into a prediction and understanding about the real world, never ceases to amaze me. And that process is largely responsible for the technological world around us.

Without physics and quantum mechanics (which is among many other things is key in the functioning of a transistor), there would be no cars, computers, Internet etc. I find it rather frustrating that people simply accept science when it makes their smart phone work, but then going to question it in terms of, for example, climate change or cosmology.

There is a somewhat different question, however, about whether studying things such as black holes, Galaxy formation, cosmology, etc., are worthwhile. Some of these do not have any practical use anytime in the near future. But the big questions — where do we come from? How does it all work? What's it all about? Those have been with humanity since the beginning. Now, we get to actually *answer* some these questions using a reliable method — even if often they just lead to bigger (and more interesting) questions. Shouldn't we?

reddit had a recent thread regarding the Universe not having a center - I've been racking my brain for two days about it and it still doesn't make sense. Maybe you could help with another or more detailed explanation?

The explanation came out on top in our thread was that the Universe is like the surface of a balloon, expanding outward but with no center - the balloon itself has a center, even if there's nothing in it, so to me that analogy only works with an artificial constraint and doesn't really *mean* anything.

Going and looking for more readings on the subject I found several that suggest the big bang is the time where the distance between the infinite points of the universe became greater than zero, therefore there's no "central point". BUT, how does that negate that the universe is expanding from a previous state of having been the smallest it was (after being nothing)...which would seem to me that you could consider a center point of the universe?

This has given me so many questions and headaches over the last two days....someone suggested the universe's center would be fourth-dimensional, unreachable by us in 3D space. If the universe is present in 3D space, does that portion of it have to follow the resident physics?

heeeeeelp!

[Azarul](#)

The balloon metaphor is good but suffers from the flaw that people always ask about the inside and outside of the balloon, when you're supposed to imagine being a flatlander stuck on its surface.

Here's another way to think about a related version of a center-less universe. Consider the ancient game Asteroids, or anything more modern where the screen wraps around in both horizontal and vertical directions. That is, as you pass off the right edge of the screen you reappear on the left. Now imagine that screen filled with little tiny galaxies placed randomly. Where is the *center* of the distribution of galaxies? You might say the center of the screen, but in terms of what each galaxy 'sees' nearby, there is no difference between the center and any other place -- the galaxies don't see the edge of the screen, since photons just go from one 'edge' to another. In other words if you translate the entire thing to the right (or left or top or bottom) you don't change anything.

Now imagine a 3-dimensional space that wraps around in just the same way -- maybe in minecraft-world instead of asteroids.

This is a model of a flat universe with a toroidal topology. A finite closed universe is technically different, but this might help conceptualize things a bit better.

What did you do for your education? Any advice to a 20 year old kid with a passion for mathematics and completely fascinated with physics?

A professor my freshman year of college told me it was too complicated to create a completely accurate model of how planets and other bodies in our solar system move. Is this true? There are many many variables to factor in, but is there some basic model that you could start with?

[inspectorallex](#)

I was lucky to go to a great public high school with an inspiring physics teacher, and to have a home with many shelves full of books on cosmology, physics, and other inspiring esoterica and two parents who supported my predilection for collecting all manner of things, building with electronics, programming (now rather pathetic) computers, and all manner of pretty geeky stuff. So being a scientist was natural for me.

I also happen to have a temperament, shared by some of my favorite colleagues like Max Tegmark, to be slightly subversive in thought but disciplined in execution, too easily bored to focus on one topic for too, too long (and too curious to stop learning about new subjects), too stubborn to give up explaining the very difficult to explain, and just skeptical enough to strategically reject the accepted explanation of many things.

My general advice would be to just keep learning, as much as you can! Getting a PhD in physics, doing a postdoc, and becoming a research scientist is a fantastic but also very challenging, competitive, and in some ways painful path. My attitude was that I would follow it as long as it kept working and going well, but be ready to abandon it -- there's lots of fun to be had in many endeavors in the world, and one must avoid the sunk-cost fallacy. So I'd advise the same.

But the learning that it takes to *understand* the really cool stuff that is happening in physics (or other fields) is much, much easier to attain. If you have a mind for math and physics, the total (full time, super-intense focused effort) time it would take to learn physics at a depth where you could understand, for example, Hawking's new paper on black hole information loss, would probably be between 6 months and 2 years. That's a long time, but it's not a 5 year PhD program.

What are your thoughts on Quantum suicide?

[lolomfgkthxbai](#)

This is a fascinating thought experiment (only -- don't try this at home!) that brings a confrontation between a subjective view of the reality and its objective description in terms of physics to a head.

As described by other posters, the idea is to become a version Schrodinger's cat: you assume that there is a quantum system that is coupled to a device that can destroy your brain instantly, with 50% probability. With the other 50% probability, nothing happens.

In the objective view, each time you pull the trigger, there result two distinct macroscopic outcomes -- one with a living person and one with a dead person. In the many worlds view of quantum mechanics, these are both equally "real". From the subjective view, until you pull the trigger these timelines lie 'on top of each other'. After you pull the trigger, you can only experience one of these timelines: the one where you live.

So if you pull the trigger again and again, you against all odds find yourself -- necessarily -- in the 'alive' system.

Now there are various things you can do with this thought experiment:

- 1) You can try to undermine its premises: don't believe the the many-worlds view, for example, or believe in some sort of immaterial, immortal soul, etc. We might imagine also that the is telling us that we are radically misunderstanding the connection between the subjective and objective views of the world.
- 2) You can make interesting variations on it. For example, suppose instead of killing you, the machine just puts you in immediate stasis for 10 hours. Subjectively, you get to pull the trigger over and over and see that the MWI is right. But after about 10 hours all of the 'yous' start waking up, and if you then ask what most of 'yous' see, they just blacked out after pulling the trigger a few times. Maybe this happens every night when you go to sleep. It just gets weirder and weirder the more you think about it.
- 3) You can worry about quantum immortality: no matter what the universe throws at us, it is only probabilistic in QM, and in some branch you will survive forever. This also gets scary fast.

I could go on all day about this one....

Thanks for the AMA!

I heard about the bubble signatures in the CMB. Whilst I'd love to think it could be the result of another universe colliding with our own, what are the other possible explanations for the signature, or the inhomogeneity in general?

[StonedPhysicist](#)

Candidate bubble collision signatures have indeed been seen by my collaborators Matt Johnson, Hiranya Peiris and others, who have looked at WMAP microwave background data. However, a careful analysis shows that these can be explained by chance. In other words, you don't see more, or more significant, bubble-like things than would be predicted by chance given the standard model of cosmology. So this place is constraints on the bubble collision model, but does not rule out the possibility that we could see something with future, more sensitive, experiments.

There are also other purported decetions of 'rings' and things in the CMB by Penrose and collaborators, but I think these are almost certainly the same effect, except they did not do a careful enough statistical analysis.

I have read, in various sources of academia, that United States funding of astrophysics research falls short in comparison to similar funding in other (mostly European) countries.

In your experience, has this proven to be true?

On another note, what are your thoughts on Ray Kurzweil's prediction about the inevitable technological Singularity occurring within the next 25 years?

[thewooba](#)

The funding situation varies from country to country and from administration to administration. Germany seems to do fairly well on a long-term basis in funding fundamental science. Italy appears to be a disaster at the moment. The UK and US fluctuate quite a bit. I strongly believe that the funding for astrophysics (and cosmology and related things) does not really reflect the public's interest in it.

Hello, Professor!

I'm studying physics and have quite a bit of interest in Shannon Entropy, information, and the like. I am thus particularly interested in the possible behaviors of Artificial Super Intelligence. My question is more on the matter of "PR" than it is on the actual physics of ASI.

Stephen Hawking (and others, but Hawking had the "biggest name") recently released the paper warning of the dangers of ASI. This has made public discussion *slightly* more possible. However, we are certainly quite a distance away from seeing politicians able to discuss the matter without sounding like they are crazy people.

From your perspective, how long/what things must happen before public discourse on preventing dangerous ASI is culturally acceptable? (in the same way public discourse about climate change is now "acceptable")

[NegativeGPA](#)

There has been a huge change in the last two years in terms of how concerns about AI and superintelligence are being heard. A few years ago this was a rather fringe way of thinking, ignored/mockered by mainstream AI researchers, and appeared almost solely in science fiction. But this year the AAAI conference and many others have had sessions, panels, etc. on AI safety and ethics. The short term issues (AI-caused unemployment, autonomous weapons, legal issues with autonomous cars, etc., etc.) are undeniable, so now that (narrow) AI is actually working in a lot of systems, almost everyone agrees that we have to talk about them.

In terms of artificial general (and probably 'super) intelligence, AI researchers are more circumspect. What almost nobody believes is that next year there will be ASI, and we'd better do something right now. But is it 10 years, 30 years, 50 years, or 500 years away. Honestly nobody knows. In my personal opinion, if it's ten years away, we should be in panic mode right now. If there's a 1-10% chance that it's ten years away, we should be very very worried. If it's 30-50 years away, we should be devoting major resources to it. I find it hard to understand anyone who could think it's 10 years away and no big deal -- we're talking about replacing ourselves as the dominant intellectual entities on the planet. Many researchers, I think, fall into the 30-50 year camp, and just want to kick the can down the road because they don't feel like we understand things well enough to do useful research. I disagree with that.

To answer your question, I think having public discussions of the short- and medium-term issues, and private discussions of the long-term issues is probably best. Discussing superintelligence "through the media" does not appear to be helpful, as every single article has a picture of the terminator on it.

1. It was a popular idea amongst futurists of cosmic era (60s-70s) to worry about alien invasion. Nowadays people are worried about threats of AI. Don't you think AI will actually be an advantage of the human race, and will actually help humanity to survive against some overevolved super smart technologically advanced aliens if (when) they invade the Earth?
2. I once read in a popular science magazine about the fact that our universe is actually a black hole, because if we calculate the Schwarzschild radius for the all seen matter of our universe, it will be bigger than the observed chunk of the universe, so we're kinda already inside a black hole. In the same article there was stated that the math behind the black hole creation implies a creation of the new universe.

So, are all universes form a meta-universe tree, and each universe creates numerous universes-children with slightly different cosmological constants?

And do you think that some of those universes have a chance to develop an intelligent life inside them, and that life evolves to the level where they can control the creation of the new universes, to increase the chance of them to create intelligent life inside them? And therefore if there evolution process amongst the self-creating universes, the intelligent life is actually a kind of evolutionary adaptation, that helps the universe to create better children?

Zolden

1. The reason to worry about human+ level AI is the same for Aliens: something with superior technology that may think very differently, and value things very differently, than we do. It's also quite possible that if there are aliens out there, they are AIs. But AI is very likely to arrive between 10 and 100 years from now, whereas I doubt aliens are. If we got a signal from space that said "We are aliens. We are vastly superior in our technology, and will be arriving in 30 years to do whatever we decide to do with Earth", we might worry about that, and we'd probably be getting busy dealing with that, no?
2. That's kind of true, in that the spacetime of a *closed* (recollapsing, big-crunching) universe is essentially the same as a the collapse of a uniform ball of stuff to form a black hole. However there is no good evidence that the universe is closed, and good evidence against the possibility of it recollapsing (because of the dark energy).
3. There is research into whether you could create a 'new universe' in a lab. The physics is speculative, to say the least, but not manifestly absurd -- i.e. you can in principle describe this using general relativity and quantum mechanics. If you could do this, then *maybe* you could figure out a way for the baby universe to inherit some but not all of the properties of the parent universe, and be even nicer for its inhabitants. This would be multiverse creationism, I guess! There are similar ideas put forward by Lee Smolin that the universe evolves so as to maximize black hole creation, because each black hole contains a baby universe.

As our horizons have ever expanded, do you think the scale of nature could ever reach a maximum limit? Like, after we discover the multiverse, maybe there's another level of multiverse, a superverse or something, then beyond that, another, then another...

D B R

The [string theory landscape](#) and the [energy landscape of folding proteins](#) are similar in many ways. So if you want to go big picture, maybe the string theory landscape underlying the multiverse is really just describing the physics on which is based the biology of some huge multiversal organism.

Put that in your pipe and spoke it!

[deleted]

[\[deleted\]](#)

Highly effective narrow AI should be a huge long-term boon, but will be a big problem in the short term. Many, many people spend most of their days doing work they do not enjoy only because they must do so to support themselves (and family etc.) Moreover a lot of the least pleasant jobs pay the worst, which has always seemed backward to me. A world in which there are only a small number of unpleasant jobs remaining, and *higher* compensation to do them, would be awesome.

But it will be tricky for us to make it into that awesome world. Although in the past increased automation has also led to the creation of new jobs to replace the old ones, there is no law of physics — or for that matter economics so far as I know — guaranteeing this. [Brynjolfsson and McAfee](#) make a pretty compelling case that we are already seeing heightening inequality resulting from recent automation, and this is very like to continue on overdrive.

I see nothing to suggest that our economic, let alone political, systems are ready for this. Such a society would require enormous wealth redistribution from the company owners to lots of people who are doing essentially nothing for that wealth in traditional economic terms. Its hard to conceive of the idea of something like that happening in the US until/unless there is huge upheaval. It seems almost guaranteed that we will go through a phase of hyper-inequality.

Moreover, AI is unlikely to just stay at the goldilocks state of doing all of our unpleasant jobs for us, without replacing us in the jobs we humans *do* want to do, and thus essentially replacing humanity altogether.

Foundational Questions Institute...

...crowd-forecasting the future.

Asimov would be proud!

Do you mind expanding a bit on how crowdsourcing predictions will work? It seems like you would get a fair bit of 'noise' or even a few factions of thought that would end up dominating the results.

[NavalMilk](#)

You can get a sense for how the interface works by just going to the site (www.metaculus.com/questions). The fun part is how to go from a collection of predictions by lots of users into a single predicted probability. Previous research shows that you can do much better than an average or median, if you do things like (a) weight predictors more if they have a strong prediction track record (which you know for each predictor), (b) discount early predictions as compared to later predictions that have incorporated more information, (c) adjust for cognitive biases -- for example when a typical person says that something is 100% certain, studies show that it tends to happen about 80% of the time!

As we gather more data, we'll be able to create a recipe for aggregating the predictions, and optimize this recipe so that it does the best possible job retroactively in getting earlier questions right. We can then use the same recipe to aggregate future questions. This will be effective as long as the number of 'knobs' in the recipe is much smaller than the training set of past predictions.

I don't know who you are but I love you for what you think.

Our government is investing billions of dollars into AI research and have said it would be used to manipulate the economy to prevent bubbles/ collapsing.

Do you think that when our government decides to do this to us, that they will also make their crime decisions/ torture tactics/ psychological manipulation decisions based on what the AI determines?

This is what I'm worried about: "These kinds of pairings have proven to yield good results on far less risky strategic battlefields. In chess games, pairing a human expert with artificial intelligence wipe the floor with teams that only use artificial intelligence or humans.

"The AIs can analyze possible moves and identify vulnerabilities or opportunities the human player might have missed, resulting in blunder-free games," Scharre explains. "The human player can manage strategy, prune AI searches to focus on the most promising areas, and manage differences between multiple AIs. The chess AI, or multiple AIs, gives feedback to the human player, who then decides what move to make." - [Former Pentagon Official Has an Idea for How We Can Avoid a Robocalypse](#)

[altheist](#)

From what I've read, a lot of thinking in the military is along these lines: keep decisions human, but use AI to dramatically increase the effectiveness of those decisions.

This could be a real boon, but share some of the same dangers of unintended consequences that other AI systems do: you might ask the AI what you should do to maximize X, but unless you really understand its plan in detail, the AI's plan for maximizing X might at the same time have all sorts of other effects you don't want and would never contemplate accepting if you knew about them. Having the AI recognize these unwanted side effects and warn you of all of them (or avoid them itself) IS the AI control problem and very difficult to solve in general.

Hello Professor Aguirre, and thank you for taking the time to answer questions on reddit.

1. The announcement from the BICEP2 project in 2014 is by many considered a mistake, since they had been too optimistic about the noise from cosmic dust. However, more nuanced interpretations of their results indicate that there still might be some substance on their findings. What do you expect from this project? Will it provide insight to your research and inflation theory in general?
2. Is it expected that projects such as advanced LIGO will be of use in determining what happened at the beginning of time (as we know it)?
3. Is your work on the arrow of time at all related to the work of [Prof. Vaccaro](#) - or is your approach fundamentally different?
4. How could we know if our universe is in a false vacuum state?

[helm](#)

1. Yes, we should certainly keep looking for B-modes caused by Gravity waves. They could still be hiding under the dust. Their discovery would be just as big of a deal as it was supposed to have been when we thought we had discovered them! It would among other things (a) be very strong evidence for inflation, (b) tell us that the energy scale at which inflation happened is very very high, (c) tell us that gravity is quantized.
2. Not so much -- advanced LIGO and perhaps someday LISA are awesome, but primarily for astrophysics rather than the early universe. There are gravitational waves that could come from

early universe sources, but most of these are not in the same detection window as LIGO and LISA.

3. Don't think so -- but I'd have to take a closer look. I think she is appealing to the particle physics T-asymmetry, and I think the arrow of time is related to the second law, and that functions largely independently of the details of particle physics.

Dr. Aguirre, how much longer do you think we have until self-sufficient AI is developed and used for war? Thank you so much for this AMA.

[Th3NXTGEN](#)

Disturbingly soon. From what I hear from AI experts like Stuart Russell, the timeline to having things like self-driving drones that could fly around looking for a particular person (or type of activity, or ethnic group, etc.) and kill them is pretty short, maybe 5-10 years.

Our military is pouring billions into AI weaponry and no doubt other countries are too. Right now, there are regulations that a human must be 'in the loop' in targeting decisions, but you can see a clear trend toward higher automation.

I don't think it's in the long-term interest of the US or anyone else to have these weapons highly developed, massively produced, and deployed all over. It's a real potential nightmare. This was the motivation behind the [open letter on autonomous weapons](#) that FLI produced, and has been signed by a huge array of AI researchers and others.

My question is about AlphaGo, the computer program that has won games of Go against the world's leading players. AlphaGo sometimes makes good moves that completely baffle its human opponents. Commentators have said that these moves may be beyond the realm of human comprehension. Do you think that AI will/does work at levels beyond human comprehension? If so, to what extent does will this hinder AI development and application? To what extent does it scare you?

Thanks for your time.

[fermats_last](#)

Yes -- certainly smart humans already do things that are beyond the comprehension of other humans who are less smart, or less skilled in a particular game, etc.

Als already do things that we can't understand, even if they are doing it worse than us -- see for example the image recognition systems that confuse a weird pattern with a starfish (<http://www.wired.com/2015/01/simple-pictures-state-art-ai-still-cant-recognize/>) I don't think we can really comprehend why this AI classifies things they way it does.

I'll get to the scary parts in another reply -- have to take a 1/2 hour break.

Can you give a best guess time frame estimate for when a general intelligence AI will be here? We already have narrow AIs and they are getting better fast, so one day we should have an AI that understands the world as a whole and subsequently it should bring the birth of a super intelligent AI.

[kwelstr](#)

I don't know. One of the motivations for [Metaculus](#) is to get a better handle on this by breaking down the problem into pieces and getting very good predictors thinking about it. I've been surprised so far that Metaculus predictors have such a short timeline for an AI passing a Turing test (see

<http://www.metaculus.com/questions/73/>)

Since I think a high-quality Turing test (say where the AI has to impersonate a PhD scientist or better a typical Redditor) is tantamount to full-blown AI, I'd be very, very worried if I agreed that there is a 62% chance of this. I don't think there is, but I may be wrong.

Hi Professor, welcome to reddit! What do you think would be the best way for someone with an undergraduate in physics to get a gentle introduction to the topics that dominate your research? What are your favourite publications?

[-Kane](#)

This one is easy ;-) a) spend time on www.fqxi.org/community b) read books like Vilenkin's Many worlds in one, Carroll's From Eternity to here, and if you dare, Penrose's Road to Reality c) Then if you want to get into the real deal, learn general relativity from Hartle's text and then Carrolls, learn quantum field theory from Zee's book, then start reading review papers from the literature.

do you truly believe that information cannot be destroyed?

[naringas](#)

This is a long discussion ;-) Although I used it myself in another post, it can be confusing to use "unitarity" and "conservation of information" interchangeably. I think it's likely that we will find a good unitary description of processes like black hole evaporation that currently lack one. I'm much less sure that there is a single unitarity description of all of the laws of physics and the universe throughout all time.

Thanks for doing this AMA.

My question is what would be the process of looking for other universes other than our own and how could we show tangible evidence to prove this.

[Chimpanzee1st](#)

The idea here is that in some versions of 'eternal inflation' (see comment above), inflation ends via the creation of 'bubbles' that form and grow. We'd be inside one of these bubbles. The bubbles can run into each other, however, and this can leave a 'bruise' on the inside of our bubble that manifests as a perturbation in the initial conditions for our observable universe. We'd be looking for those perturbations.

Hello anthany, what is your opinion on the recently discovered gravitational waves found in space about a billion light years away? Do you think this is a new direction in the research of black holes?

[pizza25](#)

My considered opinion is that it's awesome! My hope is that this discovery will provide the impetus for more and better experiments that will teach us a lot about black holes and other very interesting astrophysical objects, and maybe about ones we [haven't even thought of yet!](#)

Hey Professor I hope you're still checking in and answering questions. As someone with my dream job, how did you get to where you are today?

I am about to graduate with my BS in physics and will be applying to grad school shortly. I have a good idea of what I want to research for my concentration, but have other research interests as well. Once you're in academia and ingrained in a certain field, how difficult is it to reach out into other areas of research? From what you've posted, your research interests and other activities cover a broad area of "future technologies/questions."

Is it just a case of taking opportunities as they come? or did you have to really reach out to people in other areas for collaboration/research?

Also did you see the anisotropies and [related paper](#) from caltech's R. Chary (I'm sure there's been more published since too) that tries to provide observational evidence for your "cosmic bubble collisions"? I was pretty floored by the result, it could be in error, but Chary went through some extensive search and signal cleaning to look at the anisotropy in question. He contends there is a chance it's not a dramatic "alternate universe collision" but ends the paper saying that it more than likely IS and that even more of these anisotropies can and will be found. What do you think?

[Shelikescloth](#)

Hi,

I'm not sure how to answer your first question other than saying a lot of hard work, luck and some natural skill. The first one is what you have most control over, so work hard! Useless advice, I know...

The barriers to reaching out to other areas of research are mostly social, rather than scientific/technical. If you're an expert in one field of physics, it takes real work to learn another area enough to do research in it, but on the order of months not years. So it's an issue of being willing to spend months, which get very very hard to come by, as well as being able to 'let go' of previous research directions, which all have fun, interesting projects that you could be doing right now.

In terms of Chary's paper, it's quite interesting and I'm waiting to see how seriously the observational community takes it. The type of cosmic collision that would explain what he appears to see is somewhat different from what my collaborators and I have looked at, but it may be possible to generate some reasonable models that would explain what he sees. This is on our to-do list.

Hi - thanks for doing this.

1)As the founder of the Future of Life Institute, what scenarios absolutely keep you up at night worrying about the continued existence of the human race? Scenarios which most humans do not consider?

2) Do you think more effort should be put into sending a manned mission to Mars as a "back-up" for humanity in case of an astronomical terror wiping out the Planet?

[orchthemed](#)

1. This may sound quaint, but I'm very worry about accidental (or deliberate) nuclear war. There's a widespread sense that the danger has passed, but it hasn't. For a good scare, check out FLI's [nuclear accident near-miss timeline](#).
2. If you lived in an oasis in a vast desert, ensuring your survival by investing in protecting your oasis from huge sandstorms (maybe by making a big cave to live in) would be much more effective than going out and trying to one more dwelling out in the desert. But if you want your descendants to survive, your oasis is too vulnerable in the long term. Building one more dwelling out in the desert is

a good idea, but primarily because it's the start of building a city, or civilization of many dwellings everywhere. So at the moment I'll all for developing cheaper rocketry and starting to send stuff to mars — go Elon! But I think if we care about the future of our species we should also be investing some funds in 'reboot' projects (see, e.g. <http://fqxi.org/community/forum/topic/2061>) — and we should be investing vastly more in protecting our planet from various threats, most of which are caused by our own idiocy (see, e.g. point #1)

Go slugs!

Obligatory question. Ever done first rain?

I miss the hell out of ucsc.

[rexlibris](#)

Afraid not. Never even seen it happen -- not often on campus late at night. Plus would lead to awkwardness next day in class.

Did people fully appreciate the multiverse implications of Inflation before the erroneous BICEP2 result, or did thinking that we had 0.2 tensor-to-scalar cause people to start worrying about it?

[iorgfeffkd](#)

All of the understanding of the potential multiverse etc. was all in place prior to BICEP2 -- what BICEP2 would have done (had it actually been a detection of primordial gravity waves) would be to provide strong evidence for high-energy inflation.

Can you describe your qualifications and background regarding artificial intelligence? Why should we trust your opinion when it comes to being afraid of 'super intelligent' AI? I checked your CV (http://scipp.ucsc.edu/~aguirre/Home_files/cv_081308.pdf) but see little resources there despite your involvement with the "Future of Life Institute" -- lots of physicists, actors, and even a few graduate students in there, but seemingly no AI leadership to speak of. Seems like a critical missing piece.

[cs2501x](#)

You're right that I'm not an expert in AI research, and even if I were, you should not trust my opinion but should decide for yourself and evaluate the arguments. If you've not read 'Superintelligence' by Nick Bostrom, I'd recommend that as a very clear laying-out of the arguments. I've heard some interesting rebuttals to the arguments (and a great many uninteresting ones), but none that I have found fully persuasive, and very few that I felt fundamentally unable to evaluate due to lack of some piece of technical background.

In terms of the Future of Life Institute, Stewart Russell (AI Professor at Berkeley) and Richard Mallah (who works on AI in industry) are centrally involved, and we're constantly talking with the AI researchers in academia and industry.

Hi professor, as a high school student who wants to learn more about physics than what is taught in AP Physics, what are some resources you would recommend to learn more?

Also, what are some careers a student can set his sights on if they study physics in college as well?
Thanks in advance!

[MaybeFun](#)

Any university textbook is tailor-made for you! Pick some up that have lots of stars on Amazon and dive in!

Physics is an extremely versatile college major. It's excellent preparation for many technical fields, computing jobs, and even law school. The fundamental skill you learn in physics is actually not knowledge about physics, but rather the ability to take a problem you have never seen before, understand what tools and concepts apply, and use those to generate a mathematical, rigorous answer. That's an invaluable skill.

Hi, thank you for your time !

I was recently engaged in a discussion concerning the fine-tuned universe argument, and I was wondering what your take was on this subject ?

Also, a few years ago I heard about a theory that posits we could be able to find traces of collisions with other universes by analyzing the CMB. Have you ever heard of it, and do you think it is serious?

Finally, I have a hard time explaining what "energy" is to people. Do you have any straightforward ways to explain the concept to people with only a high school level physics education

[dospaquetes](#)

On fine-tuning: this is complex, and depends on what sort of fine-tuning argument you are thinking about.

On collisions with other universes: yes, that's the theory my collaborators and I have worked on, so I think it's serious, but also that we'd have to be very lucky to actually see something.

On energy: I guess I would say that energy is a mathematical quantity associated with a physical system that, while it can transform in the form it takes in the physical system (e.g. between heat, kinetic energy, potential energy, rest mass energy, etc.), never changes in its overall value. It is fundamental because its conservation relies only on the fact that the laws of physics don't change with time, rather than on anything about their detailed form.

Hi Anthony, Thanks for your time answering questions. It is my understanding that the following things are now true: * We have confirmed the existence of gravity waves * We are able to create momentum without ejection of mass using new EM-drive technology * At the quantum level, we have seen causality violations in recent experiments * There are serious calls for the scientific community to actively explore the manipulation of gravitational forces with electro-magnetics

About 15 years ago, I asked Steven Hawking if he believed that we would develop anti-gravity, and he answered that since we haven't seen visitors from the future, he thought it was unlikely. Now, it seems there is potential to intentionally modify local gravitational fields, perhaps only to a miniscule degree at present, which as I understand it, means we should be able to curve spacetime.

Do you believe that taken together, these advances could indicate that we may be able to experiment with, at least quantum, time travel in the not too distant future? Are you aware of any experiments that indicate any possible effects that would correspond to decoupling and recoupling a timeline in one frame of reference from another?

On the topic of AI, another that I've both thought a great deal about and also contributed to the

development of, causing me to think further... In my opinion, the key to ensuring that AI doesn't ultimately lead to the our own demise is to develop a teachable, consistent value system that would, at least initially, be involved in AI decision-making regarding action, lead to empathy or virtual but self-reinforcing empathy for humans and other organic life, and a symbiosis that would at least start AIs on a path that is friendly towards humans. The problem I keep considering is that very quickly, a true, general AI might quickly evolve to an unpredictable state that we are poorly equipped to understand, let alone continue to influence. It seems that our best chance at a long term co-existence would be to join with AI in a way that allows our consciousness to continue to be involved in AI-level decision-making. Even that, if made feasible, raises more questions than most are willing to consider today.

My biggest question to you regarding AI is how do you propose that we, as the human society we are today, work together to create a beneficent realization of AI, or at least not the nightmare scenario that many fear? Asking even more specifically, how might I get involved in these discussions beyond an AMA?

Thank you.

[miketout](#)

Hi,

I agree about the gravity waves and probably causality violations (depending on what you mean), but I'm super-skeptical about the EM-drives stuff. Would be fantastic but violating conservation of momentum is ultra-hard.

To make a time machine (or stable wormhole, or warp drive) you almost certainly need significant amounts of negative energy. That, in turn, appears to require/allow violations of the second law, which is extremely hard to do. My personal intuition is that you will never be able to accomplish using time travel anything that you could not do at least as easily by just building whatever you want to happen, and that time travel prior to the construction of a time machine is not possible.

Sorry, that's a too-short answer to a very deep question!

What, in your opinion, is the probability that a Super-Intelligent AI already exists ?

[pinosaur](#)

This is trickier than it might seem. Certainly a hand calculator is "superintelligent" in one set of tasks -- but it's very narrow. A committee or corporation or government is superintelligent (at least in principle if not always in practice), in the sense that it can make better predictions, models, decisions, etc., than an individual human mind. But a committee is not really artificial. What about Google, though? It's (at least in part) artificial, but does things that no individual human can do, partially by leveraging all of the individuals that use it, but partially using its own algorithms, processing, and data storage. Indeed Larry Page argues that Google is already a superintelligent AI, and I'd probably agree in the above sense.

What we don't have is any single computational system that by itself can do all the cognitive things a human can do, and do many of them better, that could be run at arbitrarily higher speed using more hardware, could recursively improve its own design, etc.

Hello Professor Aguirre. The metaculus.com project sounds very interesting. I've always been fascinated with any system that allows you to ask a question to a large populace and receive an answer. Can you tell us more about it, specifically how it will work?

[type_your_name_here](#)

Best is to go and see for yourself at www.metaculus.com/questions. The site is pretty new and needs new users, so if you enjoy thinking about the future and making predictions, try it out (and spread the word!)

I think Metaculus will serve some very useful purposes and it has motivations in both my work with the Foundational Questions Institute and the Future of life Institute: see blog entries I wrote [here](#) and [here](#)

You mentioned the impact of genetic engineering. Do you think that or the integration of technology into humans is more likely? And which could be more easily accepted by society? I ask because I'd give a limb to have it be replaced by a robotic limb.

[SkaForFood](#)

Like AI, genetic engineering seems to have undergone a dramatic acceleration in the past few years, especially with the confluence of low-cost genetic sequencing, big data, and the CRISPR/cas9 system.

I think there will be (as there are now) huge resources put into biotech that cures disease and illness, and that genetic engineering tools will develop quickly insofar as they can be put to this use. I expect it will be a loooooong time, on the other hand, before you'd actually want to give up your limb to get a robotic one.

About AI, probably VR intertwines with it somehow in the futuristic world. We will have people wearing VR sets pretty soon, which obviously comes with different uses, but what aspects of VR are you scared about in personal life?

[cucumberit](#)

I think Augment Reality is going to be a really, really big deal -- more so than VR in the foreseeable future. I think it is going to be quite hard to overcome the vestibular system and the human mind's very careful integration of many sensory streams into a single reality. Augmented reality can use the physical universe to supply the link, and just add things into it.

I also prefer AR in the sense that it maintains direct connections between people with each other and the world. I see my kids disappear into virtual worlds for hours at a time and feel very conflicted about it. I can see why Minecraft is in many ways more fun to build with than Legos, but I'm still curmudgeonly enough to chase them off the computer and out the door to play in the yard.

In the future games and VR games will only get more compelling and addictive. I do wonder if that's the world we want to be in.

Dr. Aguirre, What is the significance of the association (among humans) between intelligence & prosociality (e.g., <http://www.sciencedirect.com/science/article/pii/S0160289614001287>) to concerns about strong AI? Thank you.

[idster](#)

Interesting article. I'd like to believe that as beings get smarter they also get nicer -- and I hear that a lot regarding AI. But I don't see a strong basis for that belief, certainly not one that I would assign something like 99% credence to (which is what it would take for me to not worry about strong AI).

If we define intelligence as the ability to accomplish one's goals, then I see no reason to think there is a strong link between intelligence and what those goals actually are. I'd love to hear compelling

arguments to the contrary.

I am really interested in crowd-forecasting the future.

What insights have it revealed (or likely to reveal), compared to "usual" expert opinions?

[emptybucketpenis](#)

The biggest insights so far, I would say, is that it is possible: that

- a) making predictions is a skill that varies from person to person, can be increased with training, and persists over time.
- b) Aggregating predictions can result in a set of predictions that are overall more accurate than any of the individual predictors, and far better than chance.

I don't think prediction aggregation projects are yet large or developed enough that they are producing predictions that are widely used in decision-making, though I think that is coming soon, and that as projects like Metaculus start amassing large data sets there will be really, really interesting things to pull out of them. But not yet.