261: Dangerous biological research - is it worth it? (Kevin Esvelt)

Julia: Welcome to Rationally Speaking, the podcast where we explore the borderlands between reason and nonsense. I'm your host, Julia Galef, and my guest today is Kevin Esvelt.

He's a scientist at MIT, where he leads the Sculpting Evolution Group, which focuses on evolutionary and ecological engineering. He's probably most famous for inventing, in 2013, a way to alter animal populations in the wild, using CRISPR, which is a technology that basically acts like finely targeted molecular scissors that you can program to edit an organism's genome.

Our conversation has two halves. In the first half, we're talking about an important and urgent argument Kevin has been making publicly, about how research that's intended to prevent pandemics is actually putting us in a lot more danger.

In the second half, we're talking about the potential risks of Kevin's own work, on using CRISPR-based gene drives to alter animal populations in the wild. We talk about the risks and whether they outweigh the benefits, and also about just what it's like to be the guy responsible for having invented this transformative, but also potentially dangerous technology.

By the way, the audio in this conversation is a little echoey, sorry about that. The mic failed so I had to use our Zoom recording. But it's still definitely listenable.

All right, let's jump in, here is Kevin Esvelt.

Julia: Hey Kevin, thank you so much for joining me on Rationally Speaking.

Kevin: Hi Julia. Thanks so much for the invitation.

Julia: So Kevin, as I mentioned, one of the main things I was really itching to talk to you about is the debate over gain of function research. So I was hoping you could just give my audience a basic crash course in: What is gain of function research and why should we potentially be concerned about it?

Kevin: Right. Well, counterintuitively, I'm going to point out that we are concerned about gain of function research for perhaps not the right reasons and gain a function research is perhaps only a small subset of the things that we should in fact be concerned about.

Now, okay, so what is it? Well, "gain of function" implies that you are engineering something in biology to have a function that it didn't have before.

Julia: Right.

Kevin: Now this applies to pretty much everything anyone ever does in bioengineering. After all, that's much the whole point of engineering, is you build something with a function that you want. And if you're starting with something biological, presumably it didn't have that already, or you wouldn't need to engineer it in the first place. So this is why “gain of function” is a wholly inappropriate term for what we're concerned about.

What the debate is over is the question of whether scientists should attempt to make viruses that are more dangerous than the ones that we find in nature. And you can define dangerous by how transmissible they are — how likely they are to cause a pandemic. Or by how lethal they are, or otherwise likely to harm people.

But the question boils down to: Should we be trying to make viruses that are more likely to be pandemics, or more likely to be lethal, should they become a pandemic?

Julia: Okay. And is there a better term that I should use other than gain of function? What would you refer to this research as?

Kevin: That's the problem. And I think that's why they settled on the wholly inappropriate “gain of function.” We can make it a little more specific by saying, are you trying to enhance a potential pandemic pathogen?

Julia: I've heard the term dual use research, research that could be used, not just for scientific benefit, but also for harm by malicious actors. Is that also too broad in the wrong way, or...?

Kevin: Well, that's even more general. So then you're well beyond pandemic virology and into...

Julia: Colder, colder.

Kevin: ... all the different ways that you could do that. Yeah. So probably you might just say “potential pandemic enhancement,” would be... I mean here is a great question. I haven't really sat down to think what is optimal? What should we be calling this? But I would suggest it's potential pandemic enhancement.

Julia: Okay. I'll try to remember that. Feel free to correct me. So, go on.

Kevin: Yeah, so this question of, “should we make viruses that are more likely to cause pandemics?” has been the focus of debate primarily because people are, with some justification, concerned that these viruses that we make

might escape the laboratory and cause pandemics, which would be quite counterproductive.

Because all the scientists involved in this are well-meaning. It's just that some of them are most concerned about natural pandemics. And so they have this idea that if we learn which viruses might spill over from animals and cause pandemics in humans, or which ones might evolve in nature to acquire mutations that would allow them to spread in humans on a spillover, then we might be able to intervene and keep that from happening.

The folks who are concerned about accident risk say, "Well, that's all very well and good, if you can do it. But when you're working with an agent, whether in the wild tromping around a bat cave, you're running the risk of getting infected and then starting that pandemic you are hoping to prevent."

Or, once you take it back to the laboratory, we know that there is a non-trivial rate of laboratory escapes by pathogens that folks are working on in the laboratory. What's the risk that you're actually going to end up causing a pandemic due to a laboratory accident? And we actually have a ton of data on what the rate of lab leaks is. And it's not super high, but it's definitely not zero.

Julia: I wanted to get a clearer handle on what the potential risks are from this potential pandemic enhancement research... And it seemed to me that there's either two or three categories, I wasn't sure.

The first category is the accidental leak of a pathogen that has been engineered to be more transmissible or more deadly.

The second category would be something I think you haven't really mentioned yet, which is intentional misuse by a scientist, or someone who manages to get access to the laboratory and intentionally releases the pathogen.

And then the third category is someone who's not at all even affiliated with the laboratory, but who reads the papers published in academic journals about this potential pandemic enhancement research and goes, "Ah, great. Now I know how to create especially deadly or transmissible pathogens." And then they order some stuff online and they create this new pathogen and release it.

Kevin: I'm sorry, how is that different from the initial deliberate...?

Julia: So I'm just trying to ask if there's a risk even beyond people who work at that laboratory. Those are much different numbers of people — if we were

just concerned about people who work at the laboratory intentionally releasing it, then that's a small number. And maybe we can just screen them really well and monitor them. But if it's anyone in the world could create a pandemic just by reading the academic papers on this research, then that's way more worrying.

Kevin: So here in essence is the problem. The debate over pandemic enhancement research has focused almost entirely on this question of: Is it worth the *accident risk* to be able to identify pandemic viruses? Does that actually help us prevent them from spilling over into humans in the first place? Or does it give us a leg up on, say, developing vaccines against them?

And what has been almost entirely ignored is the security risk — indeed, the proliferation risk — that is inherent in identifying a virus as a credible pandemic capable pathogen. And this is newly salient… Well, it's not that new, but when SARS 1 hit, I think it is safe to say that when SARS 1 hit, the protocols required to create a coronavirus from synthetic DNA, that was capable of infecting things, had not been established. And certainly the synthetic DNA was far, far, far more expensive, such that it was really not practical to build a synthetic coronavirus when SARS 1 hit.

Today, I estimate that at least 5,000 people could create SARS 2 using the virus assembly protocols, which have been published by actually many different laboratories now. And just by ordering synthetic DNA from a supplier, using gene synthesis, you could actually have them make the entire constructs that you need, such that you don't need to know anything at all about assembling DNA. The constructs you need would come in the mail. You'd need to know and understand how to do mammalian tissue culture, and how to follow the protocol. That really is the essence of the problem — an ever growing number of individuals can make viruses.

So here's the really important part — with one exception, for which we have a vaccine already, there aren't any obvious examples of credible pandemic-capable viruses. Meaning, if someone wanted to deliberately cause another pandemic, it's not obvious where they would go.

The problem with "gain of function" research is that they are attempting to take existing animal viruses that probably can't do the job, and enhance them so that they could. And then they will share the sequence and they will publish the evidence indicating that this is a credible threat.

And as soon as they do that, we go from a world in which almost no one can make a pandemic weapon of mass destruction… And I use those words advisedly. People in the field shrink away from them. But look, SARS 2 has killed officially more than 5 million people. When you look at excess deaths, it's exceeded 15 million. That's more than any operational nuclear weapon can achieve today. So these are weapons of mass destruction.

And as soon as somebody publishes credible evidence indicating that a particular virus, of public genome sequence, would cause a pandemic... Then all of a sudden, a good 5,000 people have access to a weapon of mass destruction.

And that number will grow, just because you can do a lot of amazing things with biotech, that can really help humanity. It's a booming sector of the economy. More and more people are going into it, getting trained in these techniques. Not just PhDs, although there's actually a lot of PhDs in this. United States produces about 10,000 PhDs in life sciences broadly every year. Even if you assume that only 1% of them have the relevant skills — which I think is a massive underestimate, given how much of the focus is biomedical — then that's still a hundred extra people every year.

And that's just PhDs. And as I mentioned, some of this stuff, technicians with the right experience can do it. Certainly graduate students can do it, in many cases. So the number is just going to go up — but it's mainly that initial hump, going from "Well, nine nuclear powers to, oh, five thousand plus people."

Julia: Right. Now, is it your impression that the defenders of gain of function — sorry, of potential pandemic enhancement — research believe... Do they disagree with you about whether we should be doing and publishing this research because they think the risk is much lower than you think it is? Or do they disagree because they think the benefits are so great that they outweigh the sizable risks?

Kevin: So this is the difficult part. They have focused almost entirely on the public health consequences. That is, they are thinking of it as "Nature is the world's greatest bio terrorist. Humans are not really evil enough to misuse this stuff. Why would anyone ever do that?"

Julia: What? What?!

Kevin: "...Insofar as they might, no nation state would do this because it would hit their own people too." So security risks are just not really on their radar. They're a distraction.

Julia: So they're just... Your impression is that they're just kind of counting on every single one of the many tens of thousands of people who will eventually have access to these tools — to weapons of mass destruction, essentially — they're counting on every single one of them being rational actors.

Kevin: I think they just really haven't thought it through. So here's the problem: It's often useful to take the argument from extremes. Their hope is that by identifying... and again, I don't want to focus just on the enhancement folks, because enhancement is only part of the problem. Remember, all

that is required is a credible pandemic capable virus. It just has to be credible. It doesn't have to be made by humans. If you find one in nature and run the relevant experiments in the laboratory, that shows that it could cause a pandemic if introduced in humans, that's another one.

So there's really two groups of scientists involved in this. There's the ones who are taking, usually particularly lethal, animal viruses, that are particularly lethal, where they're known to jump in humans. And they tend to have a high mortality rate when they do. Things like H5N1 influenza, which is what caused the controversy eight or so years back when they...

Julia: Is that the avian flu?

Kevin: Yep. That's the avian flu, which they evolved to be much more transmissible in parrots, which are the best respiratory model for humans.

Julia: Didn't the scientist who published that research, just come out and say, "This is the most dangerous virus that I've ever seen"?

Kevin: Yes, that was Ron Fouchier. Which I think is actually over the top. Yeah, it would've been bad; I don't actually think that that particular variant would take off in humans.

In general, the odds of potential pandemic enhancement succeeding, to turn something that wouldn't actually take off in humans into something that would... I see it as not all that high.

It's definitely not worth it, but I'm actually more concerned by the many, many, many folks who are interested in finding every mammalian and bird virus in nature, sequencing them all, computationally analyzing them to find the ones that look most dangerous, and then performing the laboratory experiments to characterize them to determine whether or not they could cause pandemics in humans. Then, I kid you not, they intend to create a ranked order list of all these natural viruses by the potential damage that they would cause if introduced into humans

Julia: I saw someone, maybe you retweeted this or something on Twitter, say that "A good heuristic is that if your work is the kind of thing that would cause them to throw a party in a bio weapons lab, then that should be a bad sign about what you're doing."

Kevin: Yeah. That was a litmus test proposed by Ethan Alley, who's a PhD student in my group.

Julia: Yeah, that does sound like that would be a "pop the champagne" kind of list for bio weapons labs.

The thing I find so frustrating about reading the discourse on this potential pandemic enhancement research is that it seems like most defenders of the research aren't really engaging with the details of the concern that people have with it. They mostly emphasize the potential benefits of the research, the things that we could learn from it. Which may be very real, but it all seems just moot to me if the people who are concerned are right that the cost of that research is just to give tens of thousands of people the instructions to build a weapon of mass destruction.

That's the crux. And so just pointing to the benefits of the research is kind of beside the point, if you haven't resolved the concern that this research could easily kill a billion people.

Kevin: This is why...

Julia: It kind of reminds me actually — sorry. It reminds me of this web comic, maybe it was from XKCD. It was making fun of the way that some people misuse pro-con lists. It was like, "Should I drink bleach? A Pro-Con list."

And in the pro column were things like: It's inexpensive. It's low in calories. I already have some in my house.

And in the con column was just one thing: "It'll kill me."

And the caption was like, "Well, the pros outnumber the cons."

And that's sort of how I feel about this debate about the potential pandemic enhancement research. There's just the one con, but it's the only thing that matters. And you need to convince me that I'm wrong about that, that I'm wrong to worry about it, if you want me to care about the benefits.

Kevin: Well, here there's more than one con because there's the accident risk. And then there's the proliferation risk. And, I think in many ways, the proponents of pandemic enhancement have really — and pandemic discovery mind you... they've really had it good, because they're fighting the perception that the risk is from accidents.

And there, you can go back and forth. I think personally think the data on accidents suggests that the risk is sufficiently high, that it's definitely not worth it, but you know, you have to dive deep into data and argue back and forth. And of course it becomes very personal because no virologist thinks that their lab could have a leak. Very much a Lake Woebegone effect.

Julia: Right. I actually feel like a good intuition pump, kind of thought experiment, to really see if people actually believe the claim that, "Well we don't need to worry because no one would intentionally do this" is to make it even more extreme, and say, "Would you be comfortable if we gave a hun-

dred thousand people, let's just say a hundred thousand scientists, each a button that if they press it would launch a nuclear bomb. Would you feel fine because, 'Well, surely no one would do that intentionally'?"

You could even make it more analogous and say the bomb would hit their country as well, to address the argument they make that, "Well, no one would do this because it would eventually spread to their own country".

You're still kind of counting on the good intentions and rationality of every single one of those hundred thousand people. Would that not give you pause?

Kevin: Well, let's set aside the scientists who want to do this, right? Because among other things, this is their career. This is their passion. This is their vision. This is their dream. They quite justifiably want to prevent horrors like the pandemic of the last couple of years. Right? And this is how they go about it. This is their skillset. This is how they think they can help.

We're not very rational. You can't really expect everyone to weigh up all the reasons for why their career is probably net negative.

What gets me is the fact that somehow it's been framed as an issue of freedom of science. That science is under attack, if there's any restrictions on any research anywhere. And that is just very toxic because they say, "Oh, well, if you put controls on research here, then you're going to cripple the field. And we won't be able to create breakthrough cures for other things."

Julia: Right.

Kevin: "And everything is threatened if there's any restrictions on one tiny subfield of virology. That's a slippery slope, and it will lead to a corrupted scientific enterprise that misses lots of potential good that could otherwise be accomplished."

And there you have to say, "Well, okay, what do you think of physics?"

Well, what about physics? Well, is physics a corrupted weakened field? No. How about particle physics? High energy physics? No, no. Physics is a robust pillar of science. Right? We understand physics better than we understand most anything in more complex systems. Right?

How about nuclear physics? Because if there's ever been a field that's been impacted by restrictions by what they can and can not do...

Julia: I see where you're going here, yeah.

Kevin: No, they don't think that nuclear physics is a crippled field. And yet the insistence is that any kind of restrictions on virology is the beginning of the end for freedom of inquiry as academics.

Julia: Right.

Kevin: And it's even inherent in the names that cropped up in the last go around. I mean the name of the group that formed favoring pandemic enhancement was “Scientists for science.”

Julia: Wow. Way to frame the debate there. This reasoning pattern is something... I've seen stuff like this, not just from scientists who are trying to defend their own turf, or their own research, but just from people in general. From pro-science people in general, from techno optimists in general.

And what seems to me is happening is that people have a rule that they're following that's generally sound. Like “Doing scientific research makes the world better.” And then they're just not interested in potential exceptions to the rule.

Or a related example that I've seen recently is people being dismissive of concerns about existential risk, of risks that could wipe out humanity. And their dismissal is based on, “Well, doomsayers have always been wrong in the past.” Which is true. But even if you think 99% of all doomsayers are wrong, are you really going to say there's *never* going to be an exception? That it's not even possible for there to be an existential threat to humanity?

Because if you think that's possible, then you need some heuristic for making exceptions to your rule, and deciding, “Well, okay, this particular concern is plausible enough that we should investigate it.” Or, even if you're right about the general rule “scientific research is good,” you have to have some heuristic for deciding when something actually is a decent exception to that rule. And it just seems like people are not interested in considering any exceptions to the rule.

Kevin: I think that's exactly right. And to be fair to the opposition, if you look back throughout history, what year had the advance that turned out to be the worst for humanity?

And you could point to many different ones, but a decent candidate is probably leaded gasoline. I mean, it poisoned literally billions of children, cognitive stunting for generations lasting up to the present day. And yet, the same year that leaded gasoline was introduced, so was the BCG vaccine against tuberculosis.

So it's hard to point to a single year in which technological advances were net negative. It's probably been positive all the way back — so far. And we're left in this position of saying, Yep, historical evidence totally suggests that technology is always net positive and therefore we should investigate everything. We're reasoning from first principles and suggesting, well as the power of technology increases, as the accessibility of technology increases, given entropy... sooner or later, someone's going to be able to sweep their arm and knock down the city made of children's blocks. And you can't apply an opposing force to the opposite side of that city of blocks that will keep it upright when someone tries to do that.

And nuclear weapons are probably a great example of this. But we do not, in general, try to map out every possible way that one could make a nuclear weapon.

Julia: And publish it in an open journal.

Kevin: And share it with International Atomic Energy Agency inspectors so they know what to look for. And the reason we don't do that is because it's much harder for the kinds of folks who want to gain access to nuclear weapons to do that kind of research than it is for well-resourced governments who are in a position to do this. So it's actually contrary to our interests to map out the threat space surrounding the details of how to make nuclear weapons and alternative pasts thereof.

At the very least, even if we do it, we certainly don't tell others about it. Even if telling them would help inspectors do their job by preventing misuse. We still have decided, “We don't do that because we cannot defend against these things.”

But there's this strong, strong, strong intuition that if you know that there is a threat, you are always better off understanding it. And if there's no defense against that threat, then it's just not true.

And I think the last couple of years have pretty conclusively demonstrated that we really can't defend ourselves against pandemic-class threats very effectively at all. Because historically, this was a comparatively mild pandemic, at least in terms of its death toll. It was nowhere near close to anything that might even have a hope of crashing civilization.

So you hit the nail on the head in terms of the broader question, how do we persuade folks that there is a greater risk due to increasing powerful technology that could actually, not just turn one year net negative, but plausibly bring everything down? Could truly even be existential, whether due to extinction or just causing us to fall such that we can never rise, or fall into some trap that we can't escape.

None of those has ever happened before, but reasoning from first principles, we can see that they all appear quite plausible. It's just lots of people don't accept that as valid logic. They say, oh, well, that's just reasoning from first principles, empirical evidence, empiricism is against you.

Julia: Couldn't you — again, to go back to the nuclear weapons analogy — say the same thing about some malicious actor getting ahold of some nukes and using them? Surely people wouldn't look at that and say like, “Well, it hasn't happened yet, therefore we shouldn't worry about it”? That’s not considered a valid argument, so...

Kevin: But people are very reluctant to extend that one exception that's been carved out.

Julia: I see. That's very frustrating.

So another irony in all of this is that I think a lot of people, myself included, have been kind of frustrated at bioethicists for being way too conservative and risk averse in a lot of ways. Like being against human challenge trials, being against letting people volunteer to be infected with COVID so that we can develop a treatment or vaccine faster. And yet here, where we could actually use some conservatism or risk aversion, the bioethicists are nowhere to be found. It's very frustrating.

Kevin: I know of one who has even... no, two, who have even waded in to this at all. And that's like advisor-mentee pair. So partly, only-

Julia: Not independent. Right. Right.

Kevin: Not independent. And that's still on the edges of this. That's not fair, I could think of one more. But there haven't been any outspoken bioethicists on this issue. Which is just quite shocking.

Julia: It's quite surprising.

Kevin: But I do want to return to the, how we can frame this in a way that is compelling? So on the one hand, I think it's very clear that if we frame this as a question of, “Should we work with pandemics in the lab in order to prevent spillovers, versus the accident risk of doing that?” We're going to get nowhere.

Because that's the existing argument. It politicized the scientific community. You had this Cambridge working group against Scientists for Science, and the microbiology community, basically, split on it. And it's trench warfare now. It's become a partisan tribal conflict.

And so you're not going to get anywhere. You need some kind of new framing.

And so this is, well, part of why I'm raising the security argument. Because let's take the argument from extremes. They want to find every virus in nature, characterize them to determine exactly which ones could cause pandemics, and then use that to prevent all pandemics, through unspecified means.

Okay, let's assume they can do that. Let's assume they can do it a hundred percent safely so there's no accident risks, by magic. Best possible case scenario for the pandemic identification team in the current conflict: The unavoidable consequence of that scenario is that you, in fact, do have a list of viruses that are capable of causing pandemics. And therefore, you have many thousands of people who could release not one, not two, but multiple pandemic viruses simultaneously at multiple sites throughout the world, whenever they want.

Is that a good trade? To go from our world now, where we're going to get hit with, say four to five pandemics per century — each of them single virus, single site release — to a world in which thousands of people can release multiple pandemics simultaneously at multiple sites? Going from a world with nine nuclear powers to thousands of nuclear equivalent powers, with pandemic weapons of mass destruction? Is that a trade that we would make?

And I think most people look at that and would say, you know what, I think I'll pass. I'll stick with the natural pandemics.

And that's the best case scenario for pandemic identification, for pandemic prediction actually, let's call it what it is. The overarching thing that binds the “let's find all the natural pandemic capable viruses” and the enhancement folks... The enhancement folks are saying, “Well, we know that there's some that are particularly lethal. We need to figure out, are there any mutations that are accessible to them so that it might mutate and evolve to become pandemic capable.”

But they're all interested in pandemic prediction. They want to be able to predict which viruses could cause pandemics from nature, whether they exist now or whether they're likely to evolve.

And even perfect pandemic prediction, the best case scenario, giving them absolutely everything on the table, still looks net negative when you look at proliferation costs. And given that much of that cost actually accrues from the first few pandemic capable viruses you identify that are credible, it's not looking like it's one of those non-monotonic things where some

amount of pandemic prediction is good and then it turns bad later on. No, it looks like it's bad pretty much immediately.

Julia: So I guess a hole in the conversation so far has been whether there's other research that we could do, or other precautions that we could take, that would get us the same — or equivalent — benefits, in terms of preventing natural pandemics, as the researchers are pursuing with their potential pandemic enhancement research. Is there a different way to get those benefits that doesn't have all these risks and costs associated with it?

Kevin: Well, this is the deeply frustrating part. I mentioned that the, how we're going to turn this knowledge, this pandemic prediction knowledge, to something good... It's a little bit unclear actually. I mean, there's basically two broad ideas. Number one is we might be able to prevent the pandemic by teaching people to avoid the animal reservoirs that are highest risk. So just reduce human animal contact, reduce chance of spillover, reduce chance of pandemic.

Okay, but that doesn't require you to know for sure that a particular virus is capable of causing a pandemic. You can get that just by sequencing people at the interface. People who live in these environments and are likely to be exposed to things. If you just sequence them and figure out which viruses are in them, and you sequence animals in those environments and figure out which viruses are in the animals, you'll learn which animal reservoirs cause spillover risks without needing to know which particular viruses are at risk.

And so you can get most of the benefits just by doing the virus *discovery*, without doing the virus *characterization*. You don't need to predict which one, you just need to know which animals are particularly likely to be the ones to cause spillover.

And that's important because most of the folk working on this are doing the virus discovery part. And the expensive and laborious and difficult part is actually the taking it back to a lab and characterizing it, to figure out: Is it at least equivalent to a human endemic virus of the same family on all the relevant pandemic capable parameters?

So all these programs, which are using taxpayer dollars to research weapons of mass destruction, can actually continue with a very slight modification — just do the virus discovery without doing the characterization, without trying to predict individual viruses.

So that I think is sort of the plausible compromise in the end. Now, whether we should pre compromise, I don't know, that's a political question on this.

But the other benefit that they postulate is of course that it will help us in terms of mitigation measures. That we'll be able to make vaccines faster. But now we know that an mRNA vaccine can be designed within a day. And yes, you do need to know a little bit more about... So for coronavirus vaccine, we needed to know which mutations were required to stabilize a coronavirus spike protein.

So we should totally make a vaccine against one virus for every family. Doesn't have to be a human virus, can be an animal virus, but just have that experience so we know the basics of what's going on, which protein for that family we should be targeting that is particularly immunogenic, that we can generate neutralizing antibodies against. And what kinds of stabilizing mutations like that are needed in order to get it to work.

We're already doing that, thankfully. This is one of those cases where, NIH and NIAID, they're actually doing their job.

But that means that we can design the vaccine in 24 hours. And if we can produce mRNA vaccines quickly, that means we can scale them up quickly. Which means then it's a question of testing and approval. And this gets back to your point about challenge trials. We could have had vaccines approved and thoroughly tested by May of 2020, if we'd really just launched those immediately.

But that assumes that we're jumping straight to phase two. Because who in their right mind is going to do phase one trials, testing whether an mRNA delivery of a viral protein that everyone's going to see — barring a successful vaccine anyway — is safe? In practice, we're going to jump straight to phase two.

And if you jump straight to phase two, then... well, you can't run phase two until the pandemic actually hits, right? Because otherwise, what you're proposing to do is take a group of volunteers, vaccinate half of them with your experimental vaccine, and then infect all of them with a virus that has never infected humans before. And you're proposing to do this for... dozens? Hundreds? Of credible pandemic capable viruses that you've identified from nature.

That's just not plausible.

Julia: So Kevin, we've been talking about the risks and potential benefits of pandemic prediction research. And now I wanted to switch our focus to the potential risks and benefits of another kind of research that you are famous for, called gene drives, which allow us to edit the genome of an animal species.

Could you give my audience the brief intro to CRISPR-based gene drives and what we could use them for?

Kevin: Yes. So normally when we engineer an organism, we are diverting its resources to do something that we want, and we're diverting them away from its survival and reproduction. And that means that natural selection normally acts against whatever it is that we program a creature to do.

But there are some genes that don't play by the normal rules in nature. And these are genes that have found some way of copying themselves within the organism, or otherwise skewing the odds that they will be inherited when the organism does reproduce.

So you can think of CRISPR based gene drives as programming an organism to do genome editing on its own, by converting the original version of a gene to our desired engineered version, along with the capability to do that again and again and again.

So imagine instead of a normal gene, we engineer an organism in the lab, we release it into the wild, it mates with a wild version, the offspring inherit one engineered copy and one wild copy.

They mate again with another wild organism and the next generation inherit either one or zero copies. It gets diluted out because there's a lot of wild organisms out there. But if you add a CRISPR-based gene drive, then you edited the genome using CRISPR — like molecular scissors, cut the target gene, replace it with your engineered version.

But with a CRISPR-based gene drive, you all also encode the CRISPR system next to the gene that you're editing. And so when you release your CRISPR based gene drive organism, it mates with a wild type, the offspring inherit one gene drive copy, and one original copy.

But in the reproductive cells of those offspring, CRISPR turns on and it converts the original version to the engineered version, including the CRISPR system. And so when that organism mates, all of its offspring inherit one gene drive copy, and one original copy. And in the reproductive cells, CRISPR turns on and does it again, converts the original to the engineered. So if this continues, you can eventually convert the bulk of the species.

Julia: And what's an example of something we might want to use CRISPR-based gene drives to change in a species?

Kevin: Well, there's only a handful of examples where you'd want to change a whole species, but the obvious candidate is malarial mosquitoes.

If we either engineer them so they couldn't carry malaria any longer, or if we reduce their population so they're at a specialty low density that they just can't sustain malarial transmission between people, then that could be a critical component. Not a silver bullet, but a necessary if not sufficient technology for eradicating malaria once and for all.

It's a little bit hard to see how we could plausibly get rid of malaria without something like either a gene drive, a really good vaccine, or a really long lasting slow delivery therapeutic.

Julia: And roughly how many people a year does malaria kill? It's on the order of close to a million, right?

Kevin: We've gotten it down to, at least pre-pandemic, it was down to below half a million. And unfortunately that's been raising again.

Julia: That's great. That's lower than I thought, actually.

Kevin: The tragedy is most of those are children.

Julia: Right. Was it in 2014 that you discovered the potential to use CRISPR to do better gene editing?

Kevin: It was in early 2013, but I confess we sat on it for quite some time in large part because I was concerned about the implications.

Julia: So yeah, let's talk about that. How do the risks compare to the risks of the pandemic prediction research, or enhancement research, that you were talking about earlier?

Kevin: So what I eventually came to conclude is that it seems a lot like gene drive is unusual within the space of biotechnology. That is, what worried me was when I first thought of it... I was idly wondering, well, what if we wanted to edit a wild species? We'll be able to do that at scale. And including that because natural selection would act against our traits. The answer was probably not, unless you're willing to raise and release an awful lot of organisms.

And then I thought, well, what if we just program the organisms to do genome editing on their own? And then I realized that that was exactly what a gene drive system does. That's exactly the way that this yeast gene for what's called a homing endonuclease — that cuts versions of the chromosome that don't have it, and then copies itself over — that's exactly how it spreads.

And that reminded me that there was a paper that I'd read where a group of biologists at Imperial College London were working to use these homing

endonucleases from yeast to build a gene drive in malarial mosquitoes. And it wasn't working very well at all, in part because engineering homing endonucleases is a nightmare.

But I thought, well with CRISPR, not only can you easily target whatever malarial mosquito gene you want, you could target multiple sites to make it hard for the mosquito to evolve resistance by just changing the site so it can't get cut. You could engineer multiple versions of it. And in other words, you could plausibly make it evolutionarily stable.

And initially I was tremendously excited — because obviously if you could save half a million children every year, that'd be pretty amazing. And if we could do it for malaria, then why not for schistosomiasis? It's a really nasty intestinal or urinary tract worm, in effect, that causes growth stunting and cognitive stunting in children. It afflicts 250 million people, even though we have a perfect pharmaceutical cure that costs a few cents. It's just that omnipresent. But they're sexually reproducing, so we could use a gene drive to suppress the population, and plausibly drive them extinct and eradicate schistosomiasis.

So I was tremendously excited at first, but then the next morning I woke up and thought, good God. In principle, an individual researcher in the lab could just do this, just decide, we're going to engineer a whole wild species now.

Julia: Sorry to interrupt you. How big is the set of people who could just unilaterally decide to release a gene drive into some species? Is it similarly sized as the set of people who could do what I was previously calling gain-of-function research, or is it bigger or smaller?

Kevin: It's a lot fewer people than can engineer a virus, actually. And that's because engineering the germline of a multicellular organism is complicated. Because in a lot of ways, multicellular organisms evolved to prevent viruses from doing that. That is, the whole reason why we have separate reproductive system is to shield our germline, that is the DNA that we're going to pass on to our descendants, from incoming viruses and other kinds of parasitic DNA. There's all kinds of barriers in place to prevent that sort of thing from happening.

And what that means is it's really hard to get new DNA in a CRISPR system into the germline of a multicellular organism. We can do it, we've come up with ways of doing it, but it's a lot more difficult than say assembling synthetic DNA into the genome of a virus and getting it into cells in a dish.

Julia: I see. So it's a matter of skill, that's the limiting factor? Or is it just something where if you publish “Here's how to do it,” then anyone with the tools can follow your blueprint?

Kevin: So the problem right now is there are protocols for how to build a gene drive system. Designing them isn't all that difficult.

But those protocols are not going to tell you how to say, raise mosquitoes in the lab. And they're definitely not going to tell you the best technique for using a micro injector to inject mosquito eggs, nor how best to prepare those eggs, nor how best to ensure that they survive and propagate, and how to snip off one of the legs so that you can run PCR and sequence to see if you got the right kind of edit, or any of that stuff.

So there's just a lot of extra skill that is required. You've got to be a mosquito biologist in addition to knowing how to do the genome engineering.

Julia: Got it. Okay, so I interrupted you right when you were going, "Oh God, what have I done?" the morning after. So do you want to elaborate on that?

Kevin: So the "Oh God, what have I done?" was really just that, okay, not that many people could do it for any given organism, but anyone who *could* edit the genome of one organism in the lab could, in principle, edit the genomes of a large fraction of that same species — without asking any other humans for permission.

And so I spent quite some time thinking, well, what are the implications of this? And in particular, could it be misused? What if someone wanted to engineer an organism for malevolent purposes? What could we do about it?

Julia: Without giving too many creative ideas to our audience, some of whom might be scientists... What's something that someone could do intentionally to alter a species in a harmful way?

Kevin: Well, I mean the obvious example that people would give is, if we want to engineer mosquitoes so they don't spread disease, could you engineer mosquitoes so that they do? More effectively? Things like that.

Julia: I see. Diseases in addition to malaria? Or spreading malaria even more effectively than they already do?

Kevin: Things like that. Also, what about keystone species that are really important for an ecosystem to function? What if you take them out?

Julia: Or you could collapse the agricultural economy of a country by engineering their plants so that they don't grow well, or propagate.

Kevin: So no one's actually managed to build a successful gene drive in a plant yet —

Julia: Oh really? Why is that harder?

Kevin: It has to do with the frequency with which genes get copied after the DNA gets cut in plants versus other organisms, plants just don't like copying things over the way other organisms do. So it turns out we're extremely lucky in that malarial mosquitoes are unusually conducive to a CRISPR-based gene drive. It just works super well in them, which is great for getting rid of malaria.

Plants are unusually recalcitrant. And then things like other insects are something in-between. Mammals are a little bit worse than that. And one of the areas my lab is working on is ways around that, getting it to work really well in organisms where it could spread, but wouldn't be very efficient, would take a long time.

And the thing you've got to realize about gene drive is that it does take generations to spread. So a lot of people immediately say, "Oh my God, could you engineer the future of humanity by putting a gene drive in people?" And, well, you'd have to wait a long, long, long time to see the results of a gene drive in humans because our generation time is so long.

Similarly, whales, elephants, anything extremely long lived, it's just not feasible. You'd have to wait thousands of years to see any significant effect.

Julia: I guess that's another way in which we got lucky with mosquitoes, relatively speaking. Their lifespan. How long is the lifespan of a mosquito or a generation, I mean?

Kevin: Usually about two weeks.

Julia: Oh, okay, great. ... I mean, great for this purpose in particular, not great in general.

Kevin: So, getting back to this question of misuse... And I suppose it's probably worth pointing out — we've glossed over how you can suppress a population. One obvious way is, what if you engineer them to all be male?

That's one way of doing it. There's another way... so, Austin Burt actually worked out a lot of the designs in mathematics for this, long before CRISPR was a thing. Well before I was aware of it, by a decade. In 2003, he showed how you could either convert them all to males or arrange to

knock out a gene that requires only one copy for function, and showed how this could essentially spread sterility through a population.

And so we could use either of these things and CRISPR basically just lets us do it in a way that is evolutionarily stable.

Julia: Is it also possible to engineer mosquitoes so that they don't carry malaria? Is that another option?

Kevin: Yes. And so there's two competing approaches to dealing with the problem of vector borne disease. And one of them says, well, let's engineer the vector so they can't spread the disease, which has the advantage of leaving the things there in the same numbers. But it has the disadvantage of allowing the parasite the opportunity to evolve around whatever blocks you're putting in the vector using the gene drive.

Whereas if you target the vector directly with the gene drive and suppress its numbers, so there just aren't enough mosquitoes to bite someone who's infected, and then bite someone who's uninfected anymore, then that will also get rid of the parasite. But there the only way to evolve around it is for the target vector, the mosquito in this case, would have to evolve some way to resist the gene drive.

And because the gene drive uses CRISPR, we can program it to target any set of sites of our choosing. That is, it can't really out-evolve us because it's so trivial to engineer CRISPR to target whatever sequences we want. So from an evolutionary reliability perspective, suppressing the target population is probably more reliable.

In the case of malaria... this gets a little bit complicated, but basically if you've had malaria before recently, then you're somewhat resistant to getting infected again, you're unlikely to die. Which is why it primarily kills children. But if you knock malaria levels down really, really low, and then you fail to eradicate it and it comes back, then lots of people are vulnerable again. And you could end up killing a lot of people accidentally that way. So we really need to get it right.

Julia: So the thing that caused you to wake up the next morning in a cold sweat was not so much the risk of accidental harm, but the risk of intentional misuse? I guess, kind of analogous to the pandemic, engineering viruses issue.

Kevin: That's right. So I've always been a bit of an outlier in in this sense, I suppose, is that most other people who are worried about gene drive are worried the unintended side effects of editing the species to do something else. Like what happens if you knock down the levels of the main malarial mosquito, for example?

Well, I look at that and I say, you know what? There's a thousand different mosquito species in Africa. There's two of them that are really good vectors of malaria. And if we got rid of just one of them, then that would probably allow us to get rid of malaria entirely.

So one species out of a thousand seems extraordinarily unlikely to have any kind of significant ecological effects. Because if I'm a bird or a spider, I don't care which species of mosquito I'm eating, they're all pretty much the same.

Julia: This seems like such a strange disconnect. In all of the discussion of gene drives, I don't think I've... I guess I've maybe once or twice seen someone acknowledge that the set of mosquitoes that carry malaria is a small subset of all mosquitoes. But the vast majority of the discourse on this issue is just talking about the potential repercussions of getting rid of *all* mosquitoes, and the potential harm that could cause to ecosystems, et cetera, et cetera.

It's so weird to me that this very important point that you just mentioned is not more central to these debates.

Kevin: Well, maybe that's my fault for not emphasizing it at enough, I'm not sure, but that is one... I mean, there's on the order of 3,500 species in mosquito worldwide, and there's maybe half a dozen that carry human disease to any appreciable extent.

And we could suppress all of those without getting rid of the mosquitoes. We don't have to drive the mosquitoes extinct, we just need to knock them down to low enough levels that they can't transmit efficiently anymore. And then once the parasite is gone, then we can let them come back if we want.

Julia: Is it possible... I'm just trying to be charitable to the people who are worried about the ecological repercussions here. Is it possible that in the regions where there are malarial carrying mosquitoes, that in those regions, in particular, those mosquitoes are a much bigger part of the ecosystem, even though they're not that big in the grand scheme of the world?

Kevin: It's plausible that they are a comparatively larger fraction in urban areas where most of the mosquitoes feed mostly on humans. Because they tend to be the human specialist, human biting mosquitoes. But those tend to not to be the ecosystems that people are worried about when it comes to unintended consequences, urban human dominated ecosystems.

But to be fair, to steelman their argument, it's really the other side of ecological scale — that is, in the Arctic... Mosquitoes do a lot of energy trans-

fer in feeding primarily on caribou, and then getting eaten by other things at lower trophic levels.

Julia: And so if we were worried about malaria in the Arctic, then this might be a bigger deal.

Kevin: Yeah. You knock down the Arctic mosquitoes and you might actually see some ecological problems.

But simply because in the tropics, there's so bloody many different kinds of mosquitoes, that you take out one or a handful you're highly unlikely to see any changes because mostly the other species that interact with them don't care about which species, and there's enough different competing species of mosquitoes that the vacant niche is probably just going to be occupied by a substitute that's very similar.

Julia: Right. You know, even before I learned that the malarial carrying mosquitoes were a small subset of the total, it still seemed like... just back of the envelope estimate, that the potential harm, the plausible harm caused by eliminating mosquitoes was going to be small compared to the current very real harm of malaria.

I haven't done this in depth or something, but ideally I would think we would want to look at examples of other species that have gone rather suddenly extinct and look at the consequence that had. Like, I don't know, passenger pigeons or the Dodo or something like that.

My sense just from having read about this a little bit is that it's very rare for there to be a serious, longterm harm caused by eliminating one species from an ecosystem. Certainly nothing near the realm of half a million people dying every year. Do you think that's right?

Kevin: Certainly nothing near the realm of half a million people dying every year, if you're a speciesist, and value humans above others.

Julia: Right. That was kind of an implicit premise in my reasoning.

Kevin: I freely admit it. I totally am. So I am with you on that. Not everyone necessarily is, although it's true that in public debates, they're often reluctant to admit it.

But people also have this just general idea that it is... I think a key here is the action/omission distinction.

Julia: I was literally just about to bring that up!

Kevin: I hate to bring this into philosophy. But people have this idea that you're more morally responsible for the consequences of something you do, versus the consequences of something that you choose not to do.

Whereas I view them as being equivalent. So if we could use gene drive against malarial mosquitoes to save half a million lives per year, or even half of that, I would say we're morally responsible for all those deaths if we choose not to. Whereas most people would say we're responsible for whatever goes wrong with the ecosystem if we do it, and we're not responsible for the deaths if we don't. That just seems wrong to me because once you have the power, you have the responsibility. Because choosing not to use that power is also a choice.

Which does mean that as our technological power increases, so does our moral responsibility for becoming wise stewards of the planet, and all the other species. As our ability to engineer the shared environment improves, then we are morally responsible for even the suffering that wild creatures inflict on one another, if it's something that we could plausibly intervene to change.

And that's I think something that is hard for most people to accept. But you can't really blame anyone because it's really hard to see how our brains could plausibly have evolved to place moral responsibility for people choosing not to do something.

That is, I can look around and see what my neighbors are doing, and I can see the consequences of what they're doing, and if it's bad, then I can blame them for doing something that had bad consequences. But it takes a whole heck of a lot more cognitive power for me to imagine everything that my neighbors could have been doing and chose not to, because the space of possible actions is so much vastly greater than the space of things that they actually did. I just can't compute everything possible and evaluate the consequences.

So there's no way that evolution would've programmed us to do that. We of course have to use a heuristic that evaluates people on the basis of their actions because that's what we can easily observe. We can't simulate everything they could possibly have done and attach moral weight to them for not doing something when they could have easily.

With the exception being something like drowning a child where it's like, well, you were the only one within reach. That's a very obvious case where you could have pulled the child out.

Julia: Right, doesn’t this seem more analogous... It's not just some hypothetical, “maybe there's something we could be doing about malaria and we're just not doing it.” There's a specific thing that we could do about malaria, that

we're choosing not to do. That feels more analogous to the drowning child, where someone is standing there and not doing anything.

Kevin: I agree, but in my conversations with lots and lots of people from many different backgrounds and cultures, that's just not how most people see it.

Julia: I was going to say, this reminds me a little bit of the debates over the FDA and the CDC during COVID, or just in general… where a lot of people I know are very frustrated many times with the FDA for refusing to approve, or taking a long time approving something just to make sure that it doesn't cause any unintended harm.

While meanwhile, many people are dying from lack of access to the drug — but those deaths, they sort of don't count in the FDA's calculus. And I think in the calculus of many ordinary people. Because the FDA isn't directly causing those deaths, they're just failing to prevent them by approving the drug or the vaccine.

I guess I'm confused about how widespread and deep-seated that moral view is. It feels pretty alien to me, but anyway.

Kevin: Likewise. But I can definitely understand in terms of a standard heuristics point of view why people think that way. Much as I am similarly frustrated — because like, really, you can stop the clinical trial because the drug is too effective? Okay, it's unethical not to give it to the control group. Okay. So then are you going to give it to everyone? No, because it hasn't formally approved yet? But you stopped the —

Julia: I can’t, I can’t. Too many things to be angrily incredulous about the moment. So I'm just going to limit myself.

Kevin: Let's set that aside and return to, well, what could go wrong with gene drive? And what I concluded before I actually told anyone… I was a technology development fellow, not running my own lab, but I worked mostly with George Church.

And before I even told George, I sat down and thought about it in as many permutations as I could. And what I concluded is that the technology probably favors defense. My logic was the following:

CRISPR-based gene drive spreads over generations, and that means it's inherently slow. Because if you think about it, it converts at most one original copy to one engineered copy in any given generation. That is to say, it can never more than double every generation, and that's if you have perfect random mixing among mates in the population, which we never actually see.

But as an upper bound, it can never more than double every generation. And generations take time, even in short-lived species. So it's not going to spread that fast.

It is obvious, if you sequence the genome. That is, there's no way to hide a CRISPR-based gene drive. There's no way to make it subtle. If you sequence the genome, you will always see it. And the reason is CRISPR is only found in single cell bacteria and archaea. You never see a CRISPR system in a multi-cellular eukaryotic organism. But CRISPR-based gene drive only works in sexually reproducing multicellular eukaryotic organisms.

So if you sequence their genome, or even if you sequence DNA in the environment, and you see in the same sequencing read, DNA from a eukaryote and a CRISPR system, then you know a human put it there. There is no possible way to hide it.

And the Department of Defense has been studying the question of whether or not you can engineer a genome in a way that can be hidden from others. And the answer is no, you can't, because there's so many different ways that we can sequence the DNA now, that something that works for one sequencing method doesn't work on all the others.

Julia: But couldn't something spread significantly before we happen to notice that the genome had been tampered with?

Kevin: Not if we're looking for it. And so again, the same sorts of... here's where I have to confess my conflict of interest. Because I must hold myself morally responsible for the consequences of CRISPR-based gene drive, I have an interest in ensuring that we are monitoring the environment for engineered DNA that is spreading, in any form, because that would ensure that we always see a CRISPR-based gene drive early, and have time to deal with it.

And that brings us to the last consideration.

So, it's slow because it takes generations to spread, it can never more than double; it's obvious, if you sequence the genome, you can't hide it. And it's *easily countered*, that is, CRISPR allows us to cut pretty much any DNA sequence of our choice.

And what that means is: Any given gene drive system that someone else has built... I can take that, I can add additional instructions to it, telling CRISPR to cut the original version, I can engineer my version so it doesn't cut itself. And mine will continue to spread through the wild species just as effectively as the first gene drive. But whenever mine encounters theirs, mine will cut it and replace it.

And there is no way you can prevent me from doing this, because CRISPR systems are so versatile that you can't possibly engineer out all the possible target sites from your gene drive system. And what that means is that no matter what kind of change someone else makes... there are a few weird exceptions involving mode of sexual reproduction, but in general, anything malevolent you might want to do to a species, I can undo. And that's weird and different, right? The ability to perfectly counter whatever change is made just by overwriting it. It's strange and different.

So you put all these together: it's slow, it's obvious, and it's easily countered. It's really hard to make an effective weapon out of something with those characteristics.

And that's quite the opposite of, you might imagine, a pandemic. A pandemic is not particularly slow. I mean, HIV is I suppose, but most pandemics that we think of, the respiratory kind, are quite fast, especially with air travel. They are not necessarily obvious, at least SARS2 is unusually stealthy in the sense that asymptomatic transmission is a thing.

And we can't necessarily do anything about them, at least not immediately. We were lucky that the vaccines worked in this case. We still don't have a vaccine for HIV. And it looks a lot like many other things in biology do not similarly favor defense.

So in a way we're very lucky because gene drive does exist and we can point to it and say, "Hey, here is something that no one even imagined we would be able to do." At least, I have yet to find any example in science fiction or elsewhere in which anyone even hinted at the possibility that an individual researcher could change the characteristics of an entire species. It seems to have never before been imagined by any human being that that sort of world could exist.

And then, because I happened to do some early work on CRISPR, all of a sudden I saw how to do that, how to make that world our world. And that's terrifying! When the first person to imagine that a possibility exists is the same person and at the same time who sees exactly how to do it on a technical level. Because then we have no time to prepare. We have failed to imagine a possible world.

And if it had turned out otherwise, if gene drive could be effectively weaponized, we would be in deep trouble.

I mean, you can imagine a counterfactual universe where CRISPR is actually less powerful, such that you could build a gene drive system and iron out, remove all of the possible sites that your version could be targeted by another version. And then it would still be slow and obvious, but we couldn't do anything about it. And in that world... well, that counterfactual

world has a lot of problems that we thankfully don't have, that we don't have to deal with.

And that was in many ways a turning point for me because that one is around my neck, the consequences of that, for good or for ill. I say the wrong word, that screws up and delays Target Malaria by a bunch of years — that's thousands of dead kids on my head. But at the same time, if I'm wrong about some more hazardous technology, and I fail to say something there, and then it ends up being weaponized and used, then I'm responsible for those as well.

And because I sort of accidentally ended up specializing in evolutionary engineering, and specifically in the kind that can spread on its own in the wild, and because I am cursed with a security mindset in thinking about how these things could go wrong or be misused... I sort of accidentally ended up in the position of considering myself responsible for the potential misuse of biotechnology and all the possible ways that we can prevent it.

And there are, unfortunately, lots of those. But the really blindingly obvious one is, just, let's not learn to make pandemics until we can reliably defend against them. This doesn't seem hard, people. We ought to be able to handle that one.

Julia: Well, Kevin, I'll let you go finally. I really appreciate you taking all this time to come on Rationally Speaking and talk to me about pandemic prediction research, and gene drives. It's been enlightening and kind of scary.

Kevin: Well, thank you so much, Julia. It's really been a pleasure. And thank you for doing what you do and helping to preserve our world, because we need it.

Julia: That means a lot to me. Thank you, Kevin.

[musical interlude]

Julia: That was Kevin Esvelt, and if you were intrigued by this conversation you should follow Kevin on Twitter, his handle is @kesvelt. You can also read more about Kevin's work at his lab's website, Sculpting Evolution, including a nice FAQ on gene drive systems, and I'll also link to his recent op-ed in the Washington Post about the dangers of pandemic prediction research.

That's all for this episode of Rationally Speaking! Join me next time for more explorations on the borderlands between reason and nonsense.