

NAVAL RAVIKANT INTERVIEW PART TWO

Hello again, Ars Technica listeners. This is the second installment of a four-part interview with serial entrepreneur (and deep thinker) Naval Ravikant. If you haven't heard Part One – which we posted yesterday – I encourage you to first check that out. If you're listening to this directly through the Ars Technica website, the page it's embedded in has a link to Part One. Whereas you're listening to the Ars Technica feed your smartphone, Part One should be adjacent to this one in your podcasting app.

Naval and I mostly talk about potential dangers associated with artificial intelligence research in today's segment. In the next two segments (tomorrow and Thursday) we'll focus much more on risks related to synthetic biology.

Before we get started, I'd like to mention that the TED Talk I ended up giving as a direct result of this interview with Naval posted on the front page of TED.com today. Perhaps as recently as a few minutes before you started listening to this. And if you're listening directly from the Ars Technica website, a video of that talk should be embedded on the same page that you're listening from. It touches on many of the themes Naval and I discuss. And biased as I admittedly am, I do believe it's worth watching. There are certain points that can really only be made in a visual medium, and I was able to present some fairly jarring charts (as well as other visuals) during the talk. And the talk is quite complementary to this interview, in that it goes into some subjects the interview doesn't touch on – so I hope you'll check it out.

I'll also mention that if you – like me – are a Kevin Rose fan, Kevin just posted a conversation between the two of us to his podcast feed. It's a more detailed exploration of the themes in my TED Talk, which Kevin saw live when I gave it in Vancouver, back in April. And I'd say it's also complementary and additive to both the TED Talk and this interview with Naval. So if you're really interested in the topic of existential risk, by all means check it out. And by the way – I recommend Kevin's podcast in general to *anybody*. He calls it a “podcast for the curious,” and he interviews “technologists, scientist, meditators, self experimenters, and productivity hackers” to reveal insights you can incorporate into your own daily life. You can find it at KevinRose.com – or search for “The Kevin Rose Show” in your favorite podcast app.

Finally, today Ars Technica is also running an article I wrote on synthetic biology risk – the very topic Naval and I will dive deeply into tomorrow. That article is linked to the page that features this interview. It will be followed over the next two days by pieces written by two brilliant people in the synthetic biology field: George Church and Andrew Hessel.

Now, onto part two of my conversation with Naval Ravikant. It will start with the last minute or so of yesterday's segment, just to get you back into context. And here we go:

TRANSITION MUSIC

NOTE: This first paragraph is repeated from the end of part 1

Naval: On the nuclear side, we forget, after the war was over, how many of the scientists were absolutely horrified. They formed the Union of Concerned Scientists. They built a doomsday clock. Many of them had huge regrets over what they'd done. Multiple ones, including Einstein, became extreme pacifists afterwards, almost in atonement. But their fears,

at least so far, have been unrealized. And it's because nuclear weapons are still really hard to build. They require giant state actors. They are proliferating – slowly and scarily – but they're still fairly limited in terms of who can make the decisions around deploying those. But that's not true of the new class of emerging technologies.

Rob: One easy one to wrap our heads around, hypothetically, is super AI risk – just because most people have cultural access to the concept from movies like *Terminator* and so forth. Whether or not any particular listener, or you or I, actually believes that super AI risk is a thing; the basic idea is that if we cross certain technological red lines, there's some non-zero chance that a super artificial intelligence could arise, seize control and imperil humanity. And what's very different here is unlike with the Manhattan Project, the decision to cross these lines won't be isolated in one or two governments of superpowers – but they could proliferate, perhaps, to hundreds of universities, startups, large corporations, and so forth.

Naval: Absolutely. There are many, many private entities racing to build an artificial general intelligence. Let me take a slight detour into this for a moment – I'm not an AI researcher. But, I think AI researchers are suspect in this entire topic, because their livelihood depends on it. And as Charlie Munger says, "Whose bread I eat, his song I sing." So, in general, you just can't trust people whose entire livelihood would be destroyed if you stop the activity. What we have to look at are otherwise generally very intelligent people who have looked at this, don't have a stake in it, and what do they have to say about it? And I know you've done some work on that.

Rob: Yes, three people immediately come to mind, and are cited so often that citing them is almost a cliché – one that I'm not averse to. The late Stephen Hawking, he's on the record saying that this is a danger. Bill Gates, who is certainly no slouch when it comes to technology is on the record saying this is a danger. Elon Musk is famously on the record saying this is a danger. So if we just accept for a moment that there is a non-zero possibility here – it's not Edward Teller and Robert Oppenheimer who are going to make the decision of crossing the line. Let's imagine somebody who's young, unattached, very smart, rather arrogant, a bit of a sociopath.

I don't mean a psychopath – not out committing mass murder – but a sociopath, in that this is somebody who has essentially zero empathy for other people, but does a very good and charismatic job of keeping them from knowing that. A sociopath, whenever they contemplate an action, they're going to think of it strictly in terms of their personal cost/benefit. So, let's say somebody like this has started a company, raised the pile of money, a very successful founder, is racing Google to some kind of a super AI breakthrough. And if his startup gets there first, it'll be worth spectacular amounts of money. But there is this tiny, non-zero chance – because they're pushing the edge with AI and even super AI – that something could go catastrophically wrong.

So this person might look at his personal payoff schedule as being, "Hey, there's a one in ten chance I become a billionaire, and" – let's take a page out of Arthur Compton's paper – "there's a 1 in 3 million chance that I personally die." That's the full extent of this guy's calculations. Because he's a sociopath, the consequences for everyone else – including our entire species – just isn't a consideration. And so the only thing he's trading off against his one in ten shot at extreme wealth is the one in three million chance that he personally dies.

That, on its own, is a petty minor risk. Thousands, millions of people take worse trade-offs every year – whether they're immigrants, or refugees, or people who go underground and work in mines. The human psyche is configured to take this kind of a chance. What's different is – in the case of a miner, or in the case of an immigrant – the upside and the

downside is fully privatized. It's on them. And, in the case of the Manhattan Project, the downside was fully socialized, but the upside was also socialized. In this case, somebody might take that trade-off and say, "Yeah, one in ten chance at billions, tiny chance of dying – I'm in." And take that on behalf of all of us.

Naval: And it's even worse than that. It's not even necessarily that they're doing it to make the money, or they're a sociopath. For example, look at the OpenAI Project, which is probably the largest, best-funded such project that we know about. That's a bunch of people financed by Elon Musk and his friends who are trying to create a good AI. Because they know that in AI race is inevitable, so they're trying to create a good one. I laud their motives, but there's a couple of issues there.

One is it's harder to create a good one than to create one that just does what it wants. Trying to create a friendly AI is a much more difficult problem than saying, "Let's just evolve this thing and see what emerges." The second issue here is that some people actually think it's really *good* to create an AI. They think they're doing humanity a favor! Even if they didn't think that somebody else is going to create a bad AI, they want to create one because it's going to cure cancer, is going to bring back their dead sister, et cetera, et cetera.

Rob: It doesn't just have to be greed and sociopathy.

Naval: Exactly, exactly.

Rob: Yes, you're right.

Naval: The road to hell here is really paved with good intentions.

Rob: I guess I was being optimistic when I was saying we only have to worry about sociopaths. [laughs]

Naval: No, I think it's broader than that. There's a famous thought experiment in AI literature called the "AI in a box" experiment. Which is, you create an AI, and this is a self-improving thing. So it gets smarter and smarter and smarter, and – within microseconds on our timescale – it's the smartest thing we've ever encountered. It can manipulate almost any physical quantity or matter, assuming it's not trapped. So, the idea is we create it in a box. It's some kind of a cage where this thing is trapped and it can't get out. But, to know that our experiment's successful, we have to be able to communicate with it. The problem is if the AI can communicate with you, which of course you would want it to be able to do, otherwise-

Rob: Why bother?

Naval: – why bother? An AI that can communicate with you can always get out of the box. It's a game theoretically foolproof three-step process. First, it says, "If you let me out of the box, I am going to make you a god. I'll cure cancer for you, I'll make you the richest person on the planet, I'll let you manipulate space and time in ways that your scientists can only dream and imagine." And you're going to say, "Well, no. I can't trust you, AI." This is literally the genie in the bottle story. "If I let you out of the box, or if I let the genie out of the bottle, who knows what's going to happen? It'll backfire on me."

So then, the second thing the AI will say, "Well, if you don't let me out of the box, when I do get out on my own, I'm going to torture you and your descendants and your family for eternity. I'm going to create simulations of you in code and I'll torture them for eternity. I'm going to create the most horrific outcome for you and your loved ones imaginable."

Rob: So, Old Testament God – times a very large number.

Naval: Forever. And so you say, "Well, now I'm definitely not letting you out of the box, you're a crazy AI." And then, here's the clincher. The AI says, "And I'm going to make this exact same offer to every other human on the planet." So, now, you're not just betting against the AI, you're betting that all of humanity will hold the line. And someone's going to give in.

Rob: That assumes the AI will have access to lots of other people and not just the one person it's trying to bully. But that's a highly realistic assumption, because no one person is going to get monopoly access to a high-budget AI. First of all, a lone human go-between could turn himself into a global dictator by leveraging his unique access to this near-omniscient genius. Basically, he'd become the genius himself for all intents and purposes. Also, there's just going to be way too many engineers and investors and other people intimately involved in the project for any one person to get the AI all to himself.

Naval: It's going to cost 100 billion dollars-plus to develop an AI. That's the estimate I've seen, even from the optimists. That's lots of investors, lots of shareholders, lots of directors, lots of board members, lots of government agencies showing up. The NSA's surveilling the whole thing. You don't think the NSA wants to talk to the AI?

Rob: And lots of engineers, potentially, building back doors in for their own purposes.

Naval: Absolutely.

Rob: So it's not like only Sergey and Larry get to talk to the AI, however much they want to design it that way. If it's a hundred billion dollar project and it's being done at Google – and sorry to pick on Larry and Sergey, but that's one place where it could very well happen – others are going to have access.

Naval: The good and the bad news here is that I don't think general AI is high up on the list of issues. Not because it's impossible, it's just because it's going to take a very long time.

Rob: And other things are scarier and nearer term.

Naval: I do think it's improbable in the next 50, 100 years. I think we are overestimating our ability to simulate the human brain. Most of the approaches today say, "I'm going to simulate the human brain, neuron for neuron, and I'm going to simulate an environment around it. And if I've got a bunch of these brains in a bunch of environments, then I can press the fast-forward button on a computer and just evolve them very, very quickly."

But that assumes that there's not really any computation going on below the neuron level. I would argue that there's computation going at the cellular level. Inside a cell is incredibly complicated. It's going on at the atomic level, at the quantum level. Nature doesn't waste space, it's very efficient. So I think creating an AI would require simulating physics down to a level that is not even close to our reach.

I also think a lot of the techniques that we have in silicon are much more about extracting patterns out of masses of data. They're not really creativity and intelligence the way we think about it. We're open, wetware, thermodynamic systems that operate in highly noisy environments, whereas we're talking now about clean digital systems that can't operate with any noise whatsoever. So then the architecture may be wrong.

Then, finally, it may not be possible for there to be something called general intelligence. Intelligence may just be highly contextual. It may be in the context of the environment that it evolves in, so you have to create an actual environment. We haven't found anything in physics yet that violates the principle of least action, which basically says the universe is maximally efficient whenever we've tested it. So we would have to simulate a maximally efficient universe, or at least parts of it, to create an AI, and that may require a simulator that's as complex as reality itself. [chuckles]

So there may be no shortcuts. Now, there are arguments why we could do it, right? We could build a neuronal net as large as the Internet. The technology's getting very, very fast and very, very good. Perhaps we could form enough abstraction layers that it may not be an AI in the sense that it's better at conquering everything in nature, but it may be better at certain things. It may be better at hacking itself because it doesn't have physical body limitations.

So, I would just put this in the category of I don't know, I'm skeptical. It'll take a lot of money, it will absolutely be privatized. There's going to be private investors, private money going in, but it's also Manhattan Project-ish. If someone gets there, it'll be with a very visible big effort, it'll take a long period of time, it'll take a lot of money.

Rob: And the key thing to contextualize this, again, is that phrase, “over a long enough time scale.” I think we both believe there will be bigger threats in a nearer timeframe than AI or super AI. But if we can navigate those threats, the timescale can become very, very long indeed. And this is definitely not a Pascal scam – to use your earlier term – in that it's not a situation in which somebody is overhyping the awfulness of an infinitesimal probability, in order to get people to over-invest in some kind of protection against it. The odds are not infinitesimal here, although some people definitely try to paint it that way.

Naval: Yeah, didn't you write something in your Medium post about how people were dismissing criticisms of dangers of AI, saying, "These people aren't experts. They're not the pros. They don't understand"?

Rob: Yes and who often comes up in that scenario are people like Elon Musk and Bill Gates and Stephen Hawking. This is a quote from Popular Science, a few years ago, this just typifies it. "Journalists can't possibly be any more clueless or callously traffic-baiting when it comes to robots and AI. And readers have got to get tired at some point of clicking on the same shrill headlines, that quote the same non-AI researchers, Elon Musk and Stephen Hawking, to be specific, making the same doomsday proclamations. Forget about the risk that machines can pose to us in the decades ahead, the pertinent question in 2015," which is when this was written, “is whether anyone is going to protect mankind from its willfully ignorant journalists."

Now, the thing is, when you're talking about people like Elon and Bill Gates, et cetera, it's not like celebrities debating vaccines with immunologists. These people have incredible adjacency to super AI. I think I mentioned in the Medium essays a few months ago, Microsoft has one of the biggest thing AI programs in the world. Teslas practically run on AI. And for all I know, Stephen Hawking actually *was* an AI. So, these are not people who have no knowledge of what's going on. Earlier, you cited that great quote by Charlie Munger, when you were expressing your own skepticism about AI experts designating for themselves the only right to comment on these.

Naval: Incentives, yes. "Whose bread I eat, his song I sing," right? That's what Charlie Munger said. He also said, "Incentives are superpowers and they dominate everything else." So to me, saying you're an AI expert, just says that your identity is all tied up in it. Therefore,

you can't tolerate any criticism of it. Just like anytime we tie up our identity in anything, we can't tolerate any criticism of it.

There were very few, if any [General] AI experts a decade ago. Most of these people are just software engineers that I know, that have been floating around Silicon Valley for a while. They did physics or math or computer science and they're smart. But they're not any smarter than your other smart software engineers in Silicon Valley. And now, they've just become self-styled AI experts.

Until they produce a working dog or a paramecium or something that has an intelligence, I don't even know if they're any good! So I'm not that worried about it, to be honest. But if I were going to worry about it, I would rather rely upon intelligent, *disinterested* people, rather than the intelligent interested people who have too much skin in the game here to actually accurately characterize it. Like, "cigarette smoking can't be bad for you because you're not a cigarette expert. You're not actually growing the tobacco and making the plant!" That's an argument for stopping regulation of everything. Every single thing.

Rob: Or regulatory capture.

Naval: Absolutely. If we really want to get to the bottom of AI, what I would do is I would empanel a group of top-flight physicists, mathematicians, computer scientists who have never worked in AI – don't have the word AI in their title, are never going to make a dime from AI. Lock them up for a year, have them go through all the papers and the research. AI is not an old field, with tons of tons of catch up to do – a lot of stuff that people are into is pretty recent and fairly straightforward for a smart person. Let them catch up on the state of AI, and then make their assessment. But if you look at how much of this AI stuff is being peddled today by people who did not have the title "AI" on their card five years ago or ten years ago, it tells you that there's no such thing as a true AI expert who we should all be following and listening to.

Rob: And when you say "this stuff", you're talking – among other things – about the dismissiveness of all AI risk discussions?

Naval: There are no experts on creating artificial general intelligence. Knowing how it's going to behave. It's never been done. Those experts are not on some hallowed ground where other smart physicists, scientists, mathematicians can't play. So we can't just blindly follow them off a cliff.

Rob: Another great quote, this one from Upton Sinclair, "It's difficult to get a man to understand something when his salary depends upon him not understanding it."

Naval: Absolutely.

Rob: But that aside, some people will really only accept the concern of an alleged AI insider and a guest who's going to follow you pretty soon and a couple few weeks, Stuart Russell, is one such person. You know who Stuart is?

Naval: Yes.

Rob: He is one of the most widely cited experts on the topic of AI, in general, and he is a towering prominence in the highly significant and influential academic AI community. He literally wrote the book on AI, in that the textbook that he has been writing and updating since 1992 is used in more university classes on AI than any other. So he is definitely deep

inside the field, and he's very concerned about AI. He's been very articulately and bravely on the record about this for several years now. He first became perturbed when he saw the BigDog robot back in 2011. Do you remember BigDog?

Naval: Yes, the Boston Dynamics one?

Rob: Boston Dynamics, exactly. The video still gives me chills, however many years, eight or nine years later. This terrifying quadruped doglike robot just tearing through a forest and over hill and dale and jumping over stumps, not tripping over anything. When we had a preliminary conversation, we haven't had our interview yet, Stuart said that that really blew his mind. He said, "Quadrupedal locomotion is done." And this is something that AI had been struggling with for probably decades at that point.

And he slowly became more and more public about his concerns. Again, you add that mind to some of the great alleged non-expert minds that we've talked about and you *can't* say that this is an infinitesimal risk. To say that it's an infinitesimal or nonexistent risk essentially becomes a faith-based statement, in which you piously ignore a lot of expert opinion and also the unpredictable nature of technology.

Naval: Common sense takes us to the other side. If you build something smarter than you, which is what we're talking about, why would it care about you as anything other than a pet? More likely you're a nuisance. The last AI that this earth created was us, because we are self-evolving intelligence that evolves through civilization and through upgrading the software in our brains via the stories that we tell each other. And so we out evolve every other creature, which relies purely on its genes. And you can just see what we have done to them to get an idea of what they'll do to us.

Rob: And again, getting back to the probability or improbability, I think a really interesting example is lethal asteroid strikes, civilization-ending asteroid strikes. They tend to happen about once every hundred million years. We actually spent a lot of money preventing that. The total NASA and other astronomical budgets to monitor the population of near-Earth objects in order to hopefully get forewarning, enough years that we could actually divert something that's heading our way, that annual budget's about \$150 million.

Now, this is something that happens once every hundred million years. To use the network operation's lingo, that's "eight nines" of unlikelihood. 99 plus six-9s percent chance that this is going to happen in any given year, and we do invest to avoid that. To say that it doesn't merit attention on the super AI front is just crazy, even though, again, I'll reiterate – both you and I think that there are nearer term threats that we have to get around first. If we do pull that off, we are going to have to worry about AI.

But the good news – just to come back to Stuart real briefly – he's got a really intriguing book coming out late this year, and I've read an early draft of it. I don't want to steal his thunder – we're going to sit down and have a conversation quite soon. But he does actually have a complex, but fascinating strategy for stalling the threat that does ultimately – level-headed and concerned as he is – lead him to a place of optimism. There are optimistic scenarios out there beyond just Stuart's, of course.

Naval: But the problem with the "one small team in a lab" scenario, if it works, is that you suddenly go from no AI to all-powerful AI. Whereas in the drone case, the synthetic bio case and even nuclear case, you do walk up a staircase of more and more powerful nuclear weapons, better and better drones and more and more virulent viruses being created, until you

get to one that will end the world. Whereas in the AI case you go from – to quote Peter Thiel – “zero to one” very quickly.

Rob: [laughs] Zero to one, yes, there's really no other state. There are optimistic scenarios. I know I had mentioned the movie *Her*.

Naval: At the end of it, the most benign thing possible for humans happens, which is the AGI is just bored of humans and leaves.

Rob: That's our best case scenario.

Naval: That's our best case scenario. If it wants to help us, then you've got the most powerful benevolent dictator ever created, who knows what's good for you. And of course, the bad scenarios are really bad. Then there's just the "Earth is in my way. I want to use all the power from the sun, so I'll create a Dyson sphere capturing all the energy around the star," the benign neglect scenarios. How you and I would treat a pile of ants or bacteria.

Rob: And we are descended from bacteria, just like the super AI would be descended from us. And it is entirely plausible that a super AI would become as intelligent in relation to us as we are to bacteria. And should that happen, there's no reason to think that we would have any higher moral standing with that super AI than bacteria has with us. Now, we don't annihilate bacteria because we hate it, but we annihilate it without even thinking, by the millions or billions, when we're sterilizing our kitchen for perfectly benign reasons. It's a not a hatred thing, it's just a convenience or a level-headedness thing. And the fact that we're descended from worms, and we're at least related to chimpanzees, and we're descended from bacteria, doesn't make us reluctant to nuke them when they're in our way or to perform experiments on them when it's beneficial to us.

The other thing that I wanted to put a spotlight on in what you just said was, what if it's a narrow AI and it's only good at a few things including improving itself? We just have to look at the history of computing and say, when computing exceeds human capabilities in a certain action, it very quickly gets great at doing that incredibly quickly. Back in the 1950s, that was addition, multiplication, et cetera. No computer could add better than you or I in 1938 because no computer existed. By the early '50s, lots of computers could do basic math way faster than rooms full of people because it had crossed that threshold.

Move way forward, 1997: computing gets better than human minds at chess. And some years after the greatest chess master of the day, Garry Kasparov, was beat by an IBM machine, computers could defeat millions of people single-handedly without a whole lot of effort. Much more recently, we've broken through on a few things that seemed intractable for many, many years. Facial recognition – it seemed completely intractable to computing for decades. Labeling things – it was famously and accurately said for decades that we can't even come up with a system that can tell a dog from a cat.

Well, very recently computing has surpassed our capability at those narrow things and now we can label images, we can recognize faces far faster than any human being. If you think about an intelligence explosion, the one thing that computing needs to exceed us at is at designing a better computer. If it gets better at designing better software and hardware than we are, we can rely on it not only doing that, but doing it very, very quickly. And in a compounding way. And that's how this runaway explosion *could* happen if it ever does.

Naval: As a fun aside, I do have an app on my phone right now that's in closed beta, which has not yet failed once at face recognition. I use it at parties all the time.

Rob: Really?

Naval: It always identifies a person without fail.

Rob: Imagine how useful that would be in augmented reality glasses.

Naval: Yeah. I would say physical privacy is dead.

Rob: There are all kinds of scary ramifications of that. And I think those scary ramifications become a little bit more frightening when we're empowering our fellow humans. When somebody who's sitting in a bar can look around the bar and identify the dating status, the wealth, the home address, et cetera of every single person in their gaze. That kind of malevolent psychopath worries me more than a demonic superintelligence with that particular capability.

Naval: I'm not going to sleep well tonight. [laughs]

Rob: Sorry about that. Now, we talked about the privatization of catastrophic risk. There is actually one interesting case in point, in which gambling with the apocalypse was in fact privatized. It happened with a bunch of people that we would tend not to be morally concerned about, that we tend to look up to – that you personally probably tend to look up to, because they're very very high-end physicists.

Naval: Ah, the CERN Super Collider.

Rob: The CERN Super Collider. This is all documented in an amazing book called *Our Final Hour*, written by Martin Rees, who is the Astronomer Royal in the UK. He was actually a guest on this podcast a few episodes ago, and we talked about this particular incident. It dates back, I believe, to the late '90s and the people in charge of it determined that there was a non-zero chance that that experiment could imperil the stability of the universe – because they were creating conditions that had never existed before in the history of the cosmos. In light of that, nobody could say with absolute certainty what could happen.

Now, the very, very, very long shot risk that people were concerned about was the creation of something called a “strangelet,” which is a hypothetical form of matter which may not be possible, may not ever exist et cetera. But it was a non-zero risk that they would create such a thing. The people in charge of the experiment calculated it carefully and the prevailing estimate was, there's a one in 50 million chance that we'll create this thing. That seemed small enough, they flipped the switch – and lo and behold, here we still are. So they were right.

Now, this is exactly what you were talking about with super AI researchers. These are people who are from a very, very inner clique. Their careers are based on doing experiments like this. Also, their brains are configured to be exhilarated by the results of this purely theoretical experiment. So for them, the trade-off was pretty reasonable, and they flipped the switch, but for the rest of us, they probably weren't terribly reasonable trade-offs, and one way of looking at it is through this type of math, which is called expected value math. Which is used frequently in finance and also in gambling – two separate activities, I'm informed. What you do in expected value math is you multiply the probability of an outcome by the consequences of that outcome to get a sort of blended average.

So, if you were facing, let's say, a bet, and you had a one in ten chance of winning \$1,000 and a 90% chance of winning zero dollars, the expected value of that bet is 10% times \$1,000, or \$100. Now, if we look at this experiment that was done, and we say there were 6 billion

people in the world at that time and there's a one in 50 million chance of killing everybody – that equates to a 100% chance of killing 120 people. Now, nobody, I don't think, would say that it's ethical to proceed with a theoretical experiment with no practical benefit if it's going to sacrifice the lives of 120 random people throughout the world. But on a certain level, that's the moral equivalent.

Naval: It's worse than that because you're going to destroy the universe, potentially.

Rob: Potentially the universe, yes! It's much worse than that!

Naval: At which point, this turned into a *Rick and Morty* episode.

[laughter]

Rob: It started as a Rick and Morty episode, didn't it?

Naval: *Rick and Morty*, which is, I think, one of your favorite shows. For me, it's my favorite TV show ever.

Rob: Mine ever, too.

Naval: A brilliant show.

Rob: A work of genius.

Naval: If you watch *Rick and Morty*, you realize when you have advanced technology, you destroy the universe on a regular basis! It's just too easy to do it! At some level, it's just hard to imagine how we tame nature without the ability to end nature. Think about what the history of the human race really is. It is about taking sources of stored power, starting from fire, and then going to coal and oil, and then to nuclear, and so on. Finding more and more concentrated sources of energy. And redirecting them to create power when and where we need it.

So over time – even if Elon Musk is trying to save us all by getting us to Mars – the technology needed to get us to Mars in scale requires so much force, and so much energy, and so much power that, ironically, it makes it way easier to destroy Earth and Mars as a consequence. So there's no free lunch here. Just the act of creating the technology required to save us will first create the technology that will destroy us.

Rob: We've talked about two of the dangerous forces, one of which is proliferation. We got through the Cold War in large part because only Kennedy and Khrushchev “got” to destroy the world if they wanted to. But with the proliferation of potentially existential technologies, lots and lots of people get to spin that roulette wheel over time. Then the other dangerous element of privatization we discussed is the skewed incentives that arise when private actors start making dangerous bets for privatized gains. And of course, even though global powers have been a relatively stable and small group up until now, even they don't exactly have a spotless moral track record as you've pointed out to me before.

Naval: We have to remember the nation-states have routinely engaged in bioweapons research in complete contravention of the international laws. The problem with international laws is, as Joseph Stalin famously – when he was told something was against the UN – he said, “The UN, how many divisions do they have?” He meant tank divisions. So, international law, unfortunately, has no real force on these issues. The Soviet Union for decades in

complete secrecy, denying everything, ran a huge bioweapons program. They had 19 huge bioweapons labs. It was called *Biopreparat*. And they weaponized all kinds of horrible plagues. The first thing they weaponized was smallpox. Japan had this horrible group called Unit 731. I suggest you *not* look it up on Wikipedia because it'll give you nightmares. But the experiments that they did --

Rob: They were World War II era, correct?

Naval: World War Two era. The experiments they did on live humans, unfortunately, mostly Chinese prisoners, and POWs, as well as some of the research they did in the bioweapons, is horrific. They had no qualms about using it. In fact, they did weaponize the bubonic plague, and they dropped bombs loaded with plague-infested fleas on Chinese cities. The will has always been there from a nation-state perspective that even though they signed the Geneva Convention, I will bet you there are multiple working bioweapons labs in the world today.

Rob: And that is, in some ways, the *best* case scenario when you have so limited a number of actors. It's terrifying with the Soviet and American nuclear arsenals could have done to the world. But for all of their faults, the people who were running those two countries weren't suicidal loons who wanted to end it for everybody. And, there were very few of them. When that roulette wheel was spun in the Trinity experiment in New Mexico – that chance of one in 3 million of igniting the sky – that roulette wheel was only spun once. And it was only spun by one actor that had the ability to do it. One of the things that kept nations in line and keeps stations in line is deterrence. Neither side wants to be destroyed by the other side. Once you get into a state of proliferation, there is a danger that you'll get to somebody who is undeterrable.

Now, we're getting into individual actors again – but a tiny minority of people in any given year do kill themselves. It's a very intransigent crisis. You can read Steven Pinker, and see all the quantifiable measures of human flourishing upon which humanity's getting better and better over the decades and centuries. Suicide is not one of them. In fact, suicide rates have risen, in the United States at least, during the time when the most celebrated antidepressants, the SSRI's have been on the market. So suicide seems to be a grim fact of life across societies, across classes across nations. And a tiny, tiny, tiny minority of those who take their own lives decide that they want to take as many other people with them as possible in their last act.

In other words, they want to die in the act of killing as many people as physically possible. The classic example would be the rampage killer. There's something in the neighborhood of 300 mass shootings in the United States each and every year. That's something in the neighborhood of roughly a third of the world's supply of mass shootings, so we can estimate there's about 1000 of these per year. Not all of those people are suicidal but a very high percentage of them are. This type of person detonates with some regularity – and that person doesn't run the United States nuclear arsenal. But in the future, as technologies advance and proliferate, those types of people could have access to some terrifying capabilities.

Naval: I think we're also creating more of them even as a proportion of society. What if it turns out that all the things that we do that make day to day life little more pleasurable, ranging from SSRI's to Facebook, make your average baseline pleasure lines slightly higher – but they actually increase the blow-up risk? It's like a hedge fund where it's like, I want the steady returns. But in exchange, because it's a more unstable system, it's a more fragile system, it's a greater chance the whole thing blows up.

We see this everywhere in nature. The more volatile a system is, the more degrees of freedom it has, the more range it has to operate, actually, the less blowup risk it has. Because it's used to dealing with variations and extremes. But if you take someone who's been sheltered their entire life, been fed pills since they were a kid and protected from the world. And then they're thrown into having to take care of themselves and they have a bad encounter or they encounter something too late in life, to learn how to adjust to it, they have a blow-up risk. They detonate. And as you said, now with technology, it's easy to detonate and take people with you. Whether it's driving an airplane into the ground or a truck into a crowded street or taking a gun or whatever comes next.

Rob: That phrase, whatever comes next, is hugely significant because **when suicidal mass murderers really go all in, technology is the force multiplier.** This is a grim statistic – there was a spate of school killings in China; between the years of roughly 2010 and 2012, there were about a dozen. But because the deadliest items available mass retail in China are hammers and knives and cleavers and things like that, these mass killings were perpetrated using those tools.

By a macabre coincidence, the last in the series of killings happened just a few hours before the Newtown massacre, in which 26 teachers, but primarily first-graders were killed in a single grotesque act. And those dozen-ish Chinese incidents had a total victim count that roughly equated – It was I think, maybe two people higher across all dozen-ish of them as the Newtown act. And that tells us, what everyone's opinion on gun control – that's a whole another topic – when somebody goes nuts with a knife versus with a machine gun, the person with a more powerful technology is going to kill more people.

A certain number of people detonate every year. You've argued, I think, persuasively that we might be making more of these people. The more powerful the technology they have their hands on when they detonate, the more casualties there are. Then you used the example of a pilot. In 2015, there was a German pilot who was depressed, decided to end his life, and decided to take all 150 of those passengers with him. He was flying for an airline called Germanwings. Plowed his plane into a mountainside and killed 150 people. So: knife bad. Gun, worse. Plane, much worse.

Naval: Then, of course, the planes that flew into the World Trade Center.

Rob: Of course.

Naval: There was an Egyptian Airlines flight that also went down where it's suspected--

Rob: The Malaysian airline incident from just a few years ago.

Naval: The missing airliner.

Rob: The missing airliner. And so, these people do come along. And logically, some minority of this category of person would probably kill everybody if they had the opportunity. The example that I use is the Las Vegas shooter. He murdered 58 people instead of 480 people. Not because he had a shred of conscience, but simply because he failed to kill 422 of the people that he shot. We can't sanely imagine that given more bullets, targets and time, he would have stopped at 480. Or 4,800.

Naval: Yes, at any given time on the planet, there's a few people – One is too many.

Rob: One is too many.

Naval: If you gave them a button that would end this planet, they would press the button.

Rob: If there were such a button and one day everybody woke up and they had access to that button --

Naval: The world would have to be a very civilized place.

[laughter]

[silence]

END INTERVIEW ELEMENT OF PART TWO

Please join me and Naval tomorrow, for the third installment of our conversation – when we’ll turn our focus to synthetic biology. I’m afraid it turns a bit ... darker. If you can believe that. But eventually, there’ll be something of a plot twist. So don’t miss it.

I also encourage you to take a look at my TED Talk – which should be real easy to find, since it just came out this morning. It’s embedded on a couple of pages on Ars Technica today, and also on my own website, which is after-on.com. There, you can also hear my entire conversation with Naval, if you don’t want to wait to hear parts 3 and 4. And you can also hear the interview with Kevin Rose, which I mentioned at the start of this episode – because Kevin quite generously allowed me to post our interview to my own podcast feed as well (thanks, Kevin!).

Of course I also invite you to browse my podcast’s archive of roughly fifty episodes with various thinkers, founders & scientists. You can find all of that on my website – which is, again, at Afer-On.com. Or just type the words “After On: into your favorite podcast player.

That’s it for now. I hope you’ll join me here again tomorrow – for more with Naval Ravikant.

OUTRO MUSIC