

This transcript includes both Part 1 (After On Podcast #44) and Part 2 (After On #45) of the Naval Ravikant interview, *End Games*. The podcast's intro, outro, and sponsorship spots are not included in the transcripts.

Some (but by no means all) fun parts are highlighted.

## **PART ONE**

**Rob (0:06:31):** Naval thank you, first of all, for making time while we're both here in San Francisco to sit down and talk about this interesting subject – I know how busy you are. And before we dive in, I'd love to familiarize our listeners a little bit with your background, starting near the beginning. How old were you when you moved to the US?

**Naval** I was nine years old when my mother, my brother and I came to the US.

**Rob:** And where did you come from?

**Naval:** New Delhi, India. A classic American dream sort of story.

**Rob:** And when you got here – I know you've told me some stories in the past – you didn't exactly come into the top 1% lifestyle upon arrival in the US.

**Naval (0:07:04) :** I don't know if we were a bottom 1%, but we were probably close to that. So, sort of on the opposite end of the spectrum.

**Rob:** So, a long way to come. And, just to fast forward through the cliff notes of how you came from those circumstances to joining the elite ranks of Silicon Valley entrepreneurs.

**Naval (0:07:20):** I've always been a lover and student of science, and admirer of scientists. To me, growing up, scientists were the great heroes of my life. And I wanted to be a scientist – specifically an astrophysicist. Which I think is a common thing for kids of a certain age, especially boys. Back when I was growing up, you know, you get a telescope and those kinds of things. But I eventually ended up more in technology. I wanted to make money and It (technology) is the intersection of business, money, and science. And I was lucky enough, early on, to discover computers. So I fell in love with, originally, my Commodore 64, and then the Macintosh. I went to Stuyvesant, which was a magnet math and science school, which also helped in that regard. And then, by the time I got to college, it was computer science for me.

**Rob (0:08:03):** So you were a comp-sci major?

**Naval:** Comp-sci and economics.

**Rob:** Where did you go to undergrad?

**Naval:** Dartmouth.

**Rob:** And then after coming out of Dartmouth with those skills, did you come more or less directly to Silicon Valley?

**Naval (0:08:4112)** I did come directly to Silicon Valley, almost directly. I bounced through Boston briefly, at Boston Consulting Group, but I wanted to be in Silicon Valley and be in tech. And when I got here, I started working for a little ISP called @Home Network, which was doing cable modem stuff back in the first dot-com bubble. And from there, I helped start a company that made linearized optical amplifiers. I met up with some scientists coming out of Lawrence Livermore. And I was the guy who was translating their science into a business plan. And then I helped do the same thing with a 3D graphics team that eventually ended up creating what is today Google Earth. It's the group that now runs Google maps. And then I did my own dot-com which was Epinions, which went public as part of Shopping.com.

**Rob (0:08:51):** That's when I first met you, was in the Epinions days. You were four co-founders, is that right?

**Naval:** Five.

**Rob:** Five co-founders.

**Naval:** It's a pretty crowded trade.

**Rob:** It was crowded, but you guys did become part of a company that went public during the halcyon days of 2000 or thereabouts. Fast-forwarding significantly, I think you're probably best known for AngelList.

**Naval:** Currently, yes. [laughs]

**Rob:** Currently, of course. So, currently, best known for AngelList. Could you give us a quick fast forward overview of what AngelList is and what it is doing to change and disrupt the process of financing early-stage tech companies?

**Naval (0:09:23):** In 2007, I started a blog with my co-author, Babak Nivi, which was called Venture Hacks. And it was laying bare the game theory of venture capital to entrepreneurs who were raising money. We wrote a whole bunch of blog posts on how you negotiate, what term sheets look like, and so on. And the biggest question we used to get was, "Don't tell me how to negotiate a term sheet. Tell me how to get a term sheet."

I'd also raised a small venture fund, and I was investing in a whole bunch of early-stage companies, and I was an early investor in Twitter and Uber and Wish and Postmates, and a bunch of others. I basically took the list of investors that I used to work with and rather than trying to treat deal flow or company financings as this proprietary resource, I just opened it up. And I said, "I'm going to share all my deal flow with you people and feel free to share back." That's how AngelList was born.

It started as a network for entrepreneurs to meet with VCs. But eventually, we grew into a platform, where today you can do financings completely online through this product that we have called Syndicates and through angel funds that are accessible to any accredited investor. AngelList at this point has funded, literally, thousands of

companies, hundreds of millions of dollars of capital deployed. And it's the largest driver of capital into the seed stage financing marketplace.

We also run the largest recruiting site for startups at AngelList Talent, where we literally have tens of thousands of companies and over a million candidates. And we're putting hundreds of people into jobs every week into startups. That's mostly a free service. We also acquired Product Hunt, which is the place where companies go to launch their products online. So it's an ecosystem for startups to raise money, to find talent, to launch their products online.

And then last year, we spun out a business called CoinList, which is a place where you go to have your legal, regulated high-quality ICOs in crypto. I got involved a lot in cryptocurrencies a couple of years ago. I love the intersection of decentralization, the politics of it, the economics of it, the idea that we can have an Internet in the future where we don't just have a few gatekeepers who run everything. And so, I've been fascinated with crypto. I'm an amateur student of science, I read a lot of physics at night, and hang out with physicists whenever they'll tolerate me in their presence.

**Rob (0:11:25):** When did you start getting interested in the topic of existential risk, or potential catastrophic risks that humanity might face in the intermediate future?

**Naval (0:11:35):** To me, it's just an inevitable observation. When I was young, I played a lot of war games and I was an avid student of military history, which is where I learned my game theory. And so I'm very familiar with not just weapon systems, but how governments and nation-states and how individual actors behave. The trend line of weapons systems is very obvious. It's been getting easier and easier to destroy something, thanks to technology. It's getting harder and harder to defend.

And the arc of history is giving more and more power to the individual, which is great for individual liberties and freedom, but not so good when it comes to blowing things up. So it became very clear to me, probably about a decade ago, that we're just on a trend line where on a long enough time scale, I can 3D print a nuclear weapon in my house. On a long enough time scale, I can take a synthetic biology lab in my room and I can create a weaponized virus that combines the most virulent features of smallpox and the longevity of AIDS and spreads like the flu.

The technology will absolutely enable that. I can even customize it to your genetic code if I wanted to. On a long enough time scale, I think commercial air travel will come to an end because weaponized drones will be too prevalent. I think on a long enough time scale, even if nobody has any malicious intent, little Johnny could create a singularity as part of his high school physics experiment by accident. It's just the nature of the Faustian bargain with technology that we get so much more power over our natural environment. That power includes the ability to destroy things and the destructive power arrives long before the protective powers do (0:13:08 - 0:13:12).

And the protection never quite adds up, right? Look where we are in nuclear. It's much easier to destroy something with a nuclear weapon than it is to build a nuclear power plant. So I just resigned myself to it slightly fatalistically. But I assume that all

of society is already aware of this and is aware of it at scale. Culturally, science fiction has done a good job making us aware of it. But what's different now is that we're actually in range. You can actually see the timelines converging. I know you've looked at that.

**Rob (0:13:41):** Yeah, and I don't know how much of a conventional wisdom it really is. I think amongst people like you and I who have spent a certain amount of time digging into this and thinking about this stuff, everything you just said is non-controversial, but a lot of people who are further from the issues might find it very controversial. And the thing that would probably contextualize it for them better than anything else is that phrase you just used repeatedly, "Over a long enough time scale."

That's really critical because none of this stuff is necessarily going to happen tomorrow. And the unlikelihood or nigh impossibility of it happening tomorrow can train people to presume that it's never going to happen – because the present can feel very yawning and very eternal when we're inhabiting it. Are you familiar with something called Amara's law?

**Naval (0:14:25):** Yes, it's that we tend to overestimate technology in the short run and underestimate in the long run. That's partially why bubbles happen. The crypto bubble the dot-com bubble – we think it's all going to get deployed and rolled out tomorrow. We overestimate it, then it doesn't, then we're disappointed. But then a decade or two decades later, it really does change the world.

**Rob:** And it sometimes changes the world in a much grander way than we ever imagined possible even during the initial hype.

**Naval:** Right, because initially, we're just extrapolating out against known use cases. But we can't yet imagine things like – when the first dot-com came along – we can't imagine Uber, we can't imagine cryptocurrencies coming out of the internet, et cetera, et cetera.

**Rob (0:15:02):** I was thinking about it last night, and I realized a really fun example could be the history of aviation. The first manned balloon flight happened in 1783 and already two years later in 1785, you had the first person going across the English Channel. And I'd imagine sitting there looking at this in 1790 – or in 1785 – you would think that, you know, within a couple of decades, we'll probably have commercial service between Paris and London. And within a human lifetime, maybe transatlantic – and none of that ever happened.

**Naval:** Right, but they had lots of drawings of Victorian or Edwardian Zeppelins with lots of crew and passengers on board crossing all across the world.

**Rob (0:15:39):** Well, that's the kind of optimism that you would have had. I mean, in thousands of years of attempts at flight with no success – and then wow, we have somebody in a balloon and two years later they're crossing the English Channel. With that history and the inevitable disappointment it must have had, I'd imagined 120 years later if somebody said, "You know what? In a few decades, we're going to be walking on the moon." You would have thought they were insane, but a 120 years

later is when Kitty Hawk finally happened. It's after all that disappointment, all of a sudden, things really, really do take off.

**Naval (0:16:11):** Another thing here is that creating destructive power, single actor destructive power, is actually a lot easier. From the moment we first split the atom to the first nuclear bomb was a much shorter trip than to the first nuclear power plant. It always takes longer to create than it does to destroy because creativity has to be scalable. It has lots of moving parts, it has to be safe, it has to be rolled out. Whereas just blowing things up is easy. (0:16:25 - 0:16:35) So, we figured out combustion – and from there to cannonball and gun powder was much faster than to the internal combustion engine.

**Rob (0:16:42):** That makes it all the more powerful, for anybody who's skeptical when you and I start talking about some of these sci-fi improbabilities of worst case scenarios. Just think about these examples of us going from very disappointing initial progress to this explosive progress. And one thing I'd like to circle back on that you just mentioned briefly was the prediction about drones in commercial aviation. Because, by a rather cool coincidence, you first said something to me about that subject right before the Gatwick Airport incident that happened back in December.

For those who weren't aware of it – some mysterious drone activity stranded over 100,000 passengers and disrupted over a thousand flights over a span of a couple of days. You had said this audacious thing about commercial aviation and drones, I think just a few weeks before that, to me. So I'd love to drill down a little bit more about what Amara's law might have to say about drones in commercial aviation.

**Naval (0:17:38):** I've been tweeting about drones since 2013 and these are among my least popular tweets-

[laughter]

**Naval:** - because they're a little apocalyptic.

**Rob:** Least popular, in that people tweet back, "I'm really pissed off about that?" Or they just don't get liked very much?

**Naval:** They don't get liked,-

**Rob:** They don't get liked.

**Naval:** - they don't get retweeted. People just don't want to deal with it. But, at least since 2013, I've had this thought that, if you were a Bond villain trying to take over the world, the three technologies that you would look at are synthetic biology; of course, hidden nuclear weapons doomsday device style, like a Doctor Strangelove type of scenario; and hunter-killer suicide drones that are miniaturized. Nano drones that use pheromone tracking, essentially, to find their targets.

**Rob:** Pheromone tracking, talk a little bit more about that.

**Naval (0:18:16):** Essentially, it's just through your DNA signature, expressed through your breath, your breathing, your face recognition. It's not that far-fetched to see,

when the technology arrives, that you can have drones that track down heads of state and are constantly monitoring them and using that as an intimidation vector. In fact, that's what we do in the Middle East all day long. American drones are always on the lookout for Al-Qaeda or ISIS heads. The moment they stick their head out of the hole, we blow them up. And that technology is just going to get better and better and smaller and smaller. And drones are essentially just guided bullets, that's all they are. It's a bullet with a tracking system, so they will obsolete all other weapon systems; airplanes, ships, everything (0:18:43 - 0:18:52).

But getting back to your original point, I think that they easily end commercial aviation as we know it (0:18:55 - 0:18:57).

**Rob (0:18:57):** Tell me how – and by the way, how miniaturized are you thinking about? How small of a drone and how big of a swarm?

**Naval:** Well, you could go watch the drone racing videos that are online, where people race tiny drones at very high speeds with incredible maneuverability.

**Rob:** Like the size of a quarter small? Or a little bit bigger than that?

**Naval:** Maybe a little bit bigger. Maybe it's the size of your fist right now.

**Rob:** But they'll shrink so quickly.

**Naval:** They're shrinking and shrinking so quickly, yes, and their maneuverability is tremendous. That airplane that went down over the Hudson a while back that was piloted safely to the bottom-

**Rob:** The one that landed.

**Naval (0:19:26):** - and everybody survived, which was miraculous. That hit a flock of geese. So, you just need a couple of geese and they took out both engines. So, if you have a couple of geese or smaller sized drones that are relatively fast and they're just hanging out near an airport, the drones don't have to launch until the airplane takes off on the runway. And at that point, they can just take it down. So it's a terrorist wet dream.

**Rob:** Yeah, you just said it's a guided bullet. It's also a guided bullet that can hover indefinitely or it's a guided bullet that can hang out on the ground rather than in the barrel of a gun.

**Naval:** Obviously, with the gun situation, we don't want people going down to Walmart and buying an M60. Well, this is a little worse. This is going down to Walmart and buying a surface-to-air missile (0:20:02 - 0:20:04).

**Rob:** When you take the extreme case scenario of it, essentially ending commercial aviation, I'm guessing you're thinking about a proliferation scenario in which the technology becomes so cheap and trivial and the hacks that would need to be made to simple consumer, or commercial, or otherwise legal technology, would be so well understood, that effectively all airports will be shut down?

**Naval (0:20:27):** Well, Al-Qaeda and ISIS already have bomb makers. They already actually do use drones in combat, they've used them. Iran has unmanned aerial vehicles that they've been using in combat and they export a lot of technology for terrorist activities. It's not much of a stretch to see them adopting drone technology as soon as it becomes viable for attacking commercial aviation. That means that the entire radius around an airport from where you can launch a drone gets shut down, and that's a large radius.

**Rob:** The asymmetry between the destructive power and the protective technology in this particular case is pretty strongly in favor of the aggressor.

**Naval:** Drones can be guided by pure line of sight, so there's no flares you can launch. There's not even jamming that you could do effectively that wouldn't take down the aircraft itself. It just so favors the attacker asymmetrically. And in this case, unlike a suicide bomber, the attacker doesn't even have to expose themselves.

**Rob (0:21:16):** Right, they still haven't figured out who was behind the Gatwick operation. And we have to assume that that was done with very primitive technology compared to that which will be available in five years, in 10 years, in 15 years, et cetera.

**Naval:** Right, so what do you do? Do you move airports to more and more isolated locations? Well, the drones just get longer and longer range. They can fly higher and higher.

**Rob (0:21:35):** One of the things that chills me the most about the Boston Marathon attack was the fact that that was essentially the long arm of Al-Qaeda reaching out through a self-organizing cell of two unhappy brothers. And it taught them how to use a pressure cooker, something that you can find in the kitchen. It was written up in the Al-Qaeda magazine *Inspire*, or one of those online magazines. And that kind of instruction can be conveyed to somebody. You don't have to convey something that's very difficult to obtain like plutonium. You have simple instructions, that this is what you do with a pressure cooker in the case of the marathon; or this is what you do with a drone that can proliferate well, well beyond those organizations.

**Naval (0:22:12):** So, I think one thing we have to be ready for is when there's a new kind of attack on an airplane, like a shoe bomber, they shut all the flights down while they retool security to try and catch this thing. And then commercial aviation starts again 24, 48, and 72 hours later. Now, we're talking about a much more difficult scenario. Drones start taking down commercial aircraft – what do you do? There's no actual proper response.

**Rob:** Yes, and you also said something that I want to drill down on, which is a pervasive threat against heads of state. If we have a very sophisticated bad guy out there – whether it's a terrorist organization or another nation-state – that wants to basically say, "24/7 we have the capability to take out your head of state." Now, let's say we're 15 or 20 years further into the future. I could imagine that a powerful state actor, or even a non-state actor, could make the credible threat that, "Look, we've got the president more or less on our sights 24/7. And by the way, every member of Congress as well."

**Naval (0:23:03):** Or their families or their extended families or their friends. So, if you stick your neck out, essentially you're making yourself a target. George Bush and Hillary Clinton and Barack Obama pioneered droning heads of state that we were not happy with. But that technology is just going to get better and better. So our own heads of state will come under the same scrutiny (0:23:09 - 0:23:21). You know, the nightmare scenario is an automated assassination market.

**Rob:** Yeah, let's talk about this. How would that look, and how might that work?

**Naval:** It's a dystopian cyberpunk construct that's been out there for a while, so I'm not the first one to think of it, but I'll articulate it. So don't blame me for this.

[laughter]

**Rob:** We'll blame whoever came up with the idea.

**Naval:** Yeah, it's been around for a while. But assassination markets – the concept is essentially that you build a piece of code that lives peer-to-peer in the dark web and is as hard or harder to shut down than a dark market. And people can come there and they can put in small amounts of money in untraceable ways – so they could use cryptocurrencies or laundered money. And they can essentially bid up assassinations on people they don't like. So if someone's extremely unpopular – say, you're in a country where some dictator took over. All the people who fled, the diaspora, they could essentially crowdfund an assassination on this person.

**Rob:** \$5 for me, \$10 from my equally dissatisfied neighbor and eventually, you're talking real money.

**Naval (0:24:14):** Exactly. And then once that number is quite high, what happens is – depending on which dystopian scenario you're subscribed to – assassins could actually take up the bid. Or, you could have drones that are automated, almost like distributed autonomous organizations where they're running around on station and sort of self-servicing. They could be carrying out the hits. Or, the victim themselves come in and pays 2X to remove themselves from the list, which then funds the next two people.

**Rob:** Interesting.

**Naval (0:24:39):** So it becomes self-perpetuating. And given that we can't even shut down things like BitTorrent and Bitcoin and dark markets and so on, it's not clear to me you could shut down an assassination market. The way an assassin or a drone could prove that they were the ones to carry out the hit and therefore get the money, they pay a small fee and then upload the GPS coordinates and the time of when the hit is going to be carried out. And it's set on a time lock-encryption code, so it only gets decrypted after the hit was carried out. And if it matches up, then they get the money.

**Rob:** Interesting. You just answered the question that was about to pop into my mind – It's that, "How would you adjudicate this kind of thing?"

**Naval:** Smart contracts, cryptocurrencies can do it all.



**Rob:** The other thing that's scary about this is, it might just be an odious dictator with a global reputation that gathers enough money, inadvertently, to get killed. But over time, the cost of everything plummets.

**Naval (0:25:29):** Sure, it could be democratized. Essentially, you could create a piece of code that goes and downloads lists of all celebrities and automatically adds them to the assassination market list – unless they pay 2X to get off the list, which funds it further.

**Rob:** Or it might be somebody who is just unpopular at a high school. If the market-clearing cost of a hit – again, over a long and enough time scale – becomes a couple hundred dollars with cheap-ass hardware that anybody could get at tomorrow's equivalent of Radio Shack, people might be setting these things up for incredibly local unpopular people.

**Naval:** Wow, that's a very interesting point.

[laughter]

**Rob:** Sorry, I didn't mean to take it to the darkest spot, but you started it!

**Naval:** Yeah yeah, it could absolutely go there. That's the problem. Destructive technology is just spreading, and getting easier and easier. And what do we do about that?

**Rob (0:26:04):** Yeah – and when you have the asymmetry between the destructive capability and what it would take to defend against it. I mean, I can't afford personally an anti-aircraft system! **People are going to become very well-behaved, and perhaps very boring.**

**Naval:** Very quiet and very anonymous **(0:26:19 - 0:26:24).**

**Rob:** Yes, society could become a very bland place. Now, as we're starting to veer in a bit of a sci-fi direction, it's worth noting to listeners that we're not presenting spooky scenarios that we think have almost zero chance of happening. At least one, and – in almost all cases – both of us believe that each of the outcomes that we're going to talk about range somewhere between plausible, to likely, to – I think you, Naval, would say – in at least a couple of cases, almost inevitable.

**Naval:** We have to be careful of what Nick Szabo, who's the creator of BitGold and smart contracts, called Pascal scams. A Pascal scam, for people who are not familiar with it, is Blaise Pascal had this idea that you should believe in God. And the reason you should believe in God is because if you don't believe in God, and God actually exists – especially a Christian God – then you're going to spend your eternity in hell.

**Rob (0:27:13):** A lot of downside.

**Naval:** Exactly.

**Rob:** this is Pascal's wager.

**Naval:** Pascal's wager, yes. So: Infinite downside risk, minor cost to actually doing something; so you might as well do it.

**Rob:** And high upside if there turns out to be a God, and you did those small steps that Pascal advocated that you do; Now, you get eternity in Paradise, and that's pretty good.

**Naval (0:27:30):** Exactly. It's a great trade in the expected value context. Except, the problem is – well, if you want to believe in the Christian God, then you also should believe in the Muslim God for the same reason; and the Hindu Gods for the same reason; and et cetera, et cetera. There's an infinite number of those kinds of possibilities. So, eventually, you realize the human brain is just really bad at computing probabilities when infinities and infinitesimals are involved. But what's different about this is that this is not an infinitesimal probability of something with an infinitely bad outcome. This is an extreme likelihood. I would argue it's an inevitability, on a long enough time scale. And I'm not sure there is a long term here.

**Rob (0:28:08):** We differ on that. I think I'm more optimistic than you are, and that might give me a different calculus in terms of the value of fussing over it now. But a couple of things that we've just mentioned pushed me back to the same historic anecdote, which I know you know well. There was a time when there was a very small, but not infinitesimal probability that something catastrophic would happen soon after the Manhattan Project was convened.

The calculation that Edward Teller, the future father of the hydrogen bomb made, established that there was a non-zero risk that the first atmospheric test of an atomic bomb would cause a chain reaction and ignite the atmosphere. And once ignited, the atmosphere would burn away to nothing ending all life on the planet.

**Naval:** But they rolled those dice anyway.

**Rob (0:28:52):** They rolled those dice anyway. The first thing they did is they ran deeper calculations. They actually made this calculation in 1941, four years before the first test. And Edward Teller and Robert Oppenheimer and the people running the projects convinced themselves that the odds were actually zero. But not everybody agreed with that, and one person who disagreed was named Arthur Compton. He was no knucklehead – he actually won the Nobel Prize for Physics. He was in charge of creating plutonium for the Manhattan Project. He put the odds of the sky igniting at one in 3 million, and he's still considered those to be the odds when the first test went off – the Trinity test in New Mexico.

Now there's a lot of historic nuance here. But it's reasonable to say that facing this risk was a perverse public good, in that the leaders were making the choice on behalf of everyone, whether the constituents liked it or not. And none of the leaders really had to wrestle with his own greed. There was not a Manhattan Project IPO. The leaders faced substantially the same upside and downside as the general public if they bet rightly or wrongly. But as far as I know, that was the very first time that anybody had to knowingly take a very small chance of humanity being annihilated. And at that time, it was a public good.

**Naval:** Yeah, it was socialized gains, socialized losses.

**Rob:** Precisely.

**Naval (0:30:07):** As opposed to today, where if you go to Wall Street, it's privatized gains, socialized losses. Bankers make money for decades, and then every 20, 30 years, they blow up the entire financial system, costing the rest of us trillions of dollars (0:30:11 - 0:30:19).

**Rob:** Exactly.

**Naval:** You can see that same calculation now starting to arise with all kinds of private actors who have access to technology to make these same kinds of decisions where, "Do I roll the dice and end humanity?" Small probability. Or, large probability, I get famous and get rich (0:30:28 - 0:30:34).

**Rob:** And the possibility of gambling with human annihilation becoming privatized – that, again is probably going to be a counter-intuitive notion to a lot of listeners.

**Naval:** That is a very new concept. This idea of privately playing roulette with the world (0:30:44 - 0:30:47).

**Rob:** Yes, it is. To the extent that it does happen, it will be mediated by new technologies, exponential technologies that are arising right now. And it will be enabled by the asymmetric power that can be wielded by the masters of those technologies. Or if they're just in a position to take chances that the rest of us aren't.

**Naval (0:31:06):** 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 (0:32:26):** 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 (0:32:59):** 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 (0:35:28):** 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 (0:36:17):** 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 (0:36:27):** 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 (0:37:00):** - 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 (0:37:55):** 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 (0:38:54):** 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 (0:39:37):** 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 (0:41:50):** 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 (0:42:37):** 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 (0:43:57):** 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 (0:45:06):** 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 (0:45:57):** 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 (0:46:36):** 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 (0:47:15):** 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 (0:48:12):** 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 (0:48:21 - 0:48:38).

**Rob (0:48:38):** 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 (0:50:04):** 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 (0:50:45):** 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 (0:51:07):** 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 (0:53:36):** 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 (0:54:00 - 0:54:08). 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 (0:54:19):** 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 (0:54:42):** 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 (0:57:13):** 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 (0:57:32):** 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 (0:57:41 - 0:57:45). 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 (0:58:21 - 0:58:28).

**Rob (0:58:28):** 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 (0:59:11):** 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 (0:59:58):** 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 (1:00:27):** 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 (1:02:33):** 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 (1:03:36):** That phrase, whatever comes next, is hugely significant because when suicidal mass murderers really go all in, technology is the force multiplier (1:03:41 - 1:03:47). 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 (1:05:11):** 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 (1:05:59):** 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] (1:05:53 - 1:06:11)

.

## **PART TWO**

**Naval (02:06):** 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 (02:21):** The world would have to be a very civilized place. You would basically have to make everybody really happy. But that's what we're talking about, over a long enough timescale everyone's going to have that button. Again: long enough timescale, that's where we're headed. Going back to the physicists for a second, one of our favorites, Enrico Fermi is well known for the Fermi Paradox. The Fermi paradox is, if the universe is so large and it's so old, how come we haven't encountered any other intelligent life yet?

For a little while, we had this very parochial point of view, which seemed like an offshoot of the hypothesis that the earth was the center of the universe before Tycho Brahe and Johannes Kepler disabused us of that. People used to have this notion, the world is unique, the earth is special, there's only one planet with water and with life. But now, we've found thousands, tens of thousands of planets, as we launch bigger and bigger telescopes, that are capable of bearing life.

So the universe is littered with them and we have only been around this civilization for 5 or 7,000 years. We've only been around as a species for a few hundred thousand years. We've only been around as life on this planet for a few billion years. Where is everybody? I actually think the most likely, plausible explanation is the



Great Filter hypothesis – and the Great Filter hypothesis is that any sufficiently advanced civilization, blows itself up.

**Rob (03:36):** When people started first wrestling with that possibility – back, probably, in the '50s, when the Drake equation was first created, that asked how many advanced civilizations might be out there – they were thinking in terms of nuclear danger. But today we're seeing other technologies that are on the intermediate, even near-term horizon that can also potentially have an annihilating impact.

**Naval (03:56):** Yes, before we even get to our favorite topics synbio which is, I think, where much of the near term danger comes from; on the Great Filter hypothesis, there's another one that was recently popularised by I believe Cixin Liu. Apologies if I get his name wrong, but the famous sci-fi author who wrote the *Three Body-Problem*. He has a hypothesis called the Dark Forest hypothesis, where the universe is like a dark forest that we're all tiptoeing through. And whoever makes a noise identifies themselves to the other predators in the forest, who immediately kill it. Because we all have killing power and first strikes are so powerful in a technologically advanced society, that whoever finds the other one first, instead of saying hello, they just wipe you out, just in case. It's the exact opposite of the prime directive from *Star Trek*.

[laughter]

The problem with the Dark Forest hypothesis is that any species that's capable of destroying another species immediately on contact was first capable of destroying itself. So how did it make it through that whole situation? Through that great filter? The only solution that I can come up with is not a great one, but it's the Borg solution which is, you all become one thing. I don't try to kill myself. You don't try to kill yourself. Even though individual cells may go rogue, they don't have the ability to shut your whole body down, because we've decided the unit is one thing and has a central intelligence and a central command.

**Rob (05:11):** Is the presumption that Borg-like intelligence would, in an abundance of caution, annihilate any species that they came across?

**Naval:** On that one, I'm a little optimistic, because I think to become a Borg-like civilization you would have to have become pacifists first.

**Rob:** In order to survive to that level.

**Naval:** That's right. You would care more about the collective.

**Rob (05:28):** Well, you had mentioned synbio. To me, the powerful example is the Human Genome Project. Which is really, technically, a step before synthetic biology – it's about reading DNA matter rather than altering it. But there's something that is sometimes referred to as the Carlson curve, which is the curve that traces how rapidly the price and performance improves in the realm of genomic sequencing and synthetic biology in general. And it makes the Moore's Law curve of computing look pokey. The statistic is, Human Genome Project ended in 2003. It took 13 years, it cost about \$3 billion. It involved thousands of the top scientists throughout the world, and its result was a single haploid genome.

And today, substantially more work can be achieved for about \$1,000. So, one three millionth the cost of the Human Genome Project. By a lone lab tech – which might very well be an undergrad or even a high school kid – working with the proper equipment in the span of about a day or maybe two days. That's the rate of improvement that we have.

And, in contemplating what might become possible with synthetic biology – particularly viewing it through this grim lens that we're looking at things through right now – try to imagine the field of synthetic biology today. Everything that it's doing that's amazing with CRISPR and with other technologies. Imagine the total output of that field over the coming 15 years. Who knows what it will be? But we can imagine it will be profound. And then, imagine that unit of work – that which the entire field achieves in 15 years – eventually becoming an easy days work for a lone undergraduate. Would that be a more or less stable world than a world of thousands of sovereign nuclear powers? Because if we had gotten as good and efficient with nukes after the Manhattan Project, you would have had nuclear weapons in thousands of garages and college laboratories by the mid 1960s.

**Naval (07:18):** Synthetic biology will be the first one to deliver world-ending capability to individuals. (07:18 - 07:24)

**Rob:** It'll come before nanotechnology.

**Naval:** Yes, nanotech is still way off.

**Rob:** Other technologies that sometimes people site as potentially having very asymmetric powers in the future include geoengineering.

**Naval:** There's always the possibility for a black swan. By definition, they're unpredictable. But the clear, current progress rate of synthetic bio may have already gotten us there. You had this horrific example you gave about how they modified one of the flu viruses.

**Rob (07:49):** Yeah, and this was in the ancient year of 2011. So this is the kind of thing that not only *could* be done, but *has* been done. It has been done with tools that are, by today's standards, relatively primitive. Back in 2011, researchers in Holland and Wisconsin, were working with something called H5N1 flu. Now, H5N1 flu is literally 3,000 times more lethal than swine flu. And swine flu is very rightly feared. But H5N1 kills 60% of the people that it infects, whereas swine flu kills 0.02% of the people it infects – so, 3,000 times more lethal.

The one thing we can like about H5N1 flu is, it is not very contagious at all. There's only been, I believe, a few dozen deaths from H5N1 in the entire history of the world. What these researchers did is they created a highly contagious version of H5N1. And they kept it under lock and key and they did it as part of a virology experiment. These weren't bad guys. But the Journal *Science*, writing about this experiment – and *Science* is one of the two most revered publications in the world of scientific research, the other being *Nature*. The news arm of the journal *Science* wrote, this strain could “Change world history if it were set free” by triggering a pandemic “quite possibly with many millions of deaths.”

At the same time, there was a person whose title was the US National Science Advisory Board Chairman for biosecurity – his name is Paul Kime. He said, "I can't think of another pathogenic organism that is as scary as this one. I don't think anthrax is scary at all compared to this." And this guy was an anthrax expert. So that happened back in 2011. Now, there's an optimistic way of viewing this, which is that this was a very carefully sequestered critter. It was probably deep in some biosafety vault on both sides of the Atlantic. And there was no way it was going to teleport into your Caesar salad tomorrow.

Another way of looking at it is they created, basically, very tiny data files. The whole flu genome is about 15,000 letters. That's very, very small. The number of modifications that went into this particular mod, I don't know, but my guess is they could probably fit on a Post-It note. And anybody who gets their hands on that tiny list of changes – whether it's in 2011, today or 30 years in the future – would have a trivially easy time reanimating that particularly contagious strain of H5N1.

Now, 2011 was before CRISPR was invented. CRISPR is an editing technology, which has radically enabled and radically eased the process of editing a genome. It has diffused to the point where CRISPR is widely practiced in high-end, high school biology labs. And I'd say that's a good thing. I think synthetic biology has an enormous amount of promise for humans, and it should be taught to young folks, et cetera. But there is this terrible danger that as this proliferates, there will be more and more turbocharged pathogens generated.

**Naval (10:47):** What we've basically done with synthetic biology, we've taken designing and hacking humans from the physical domain to the digital domain (10:49 - 10:55).

It's a similar problem that people keep creating new computer viruses. But unlike computers, when a human crashes, it's fatal. There's no reboot (10:59 - 11:03).

**Rob:** And I think it's safe to guess that a desktop DNA synthesizer from the year 2038 will be capable of things that the entire project of synthetic biology is incapable of today. There's a writer named Richard Preston, who wrote a book that some people probably know called *The Hot Zone*. He's written fiction that's based on bioterror and bio-error; he's written science fact that's based on bioterror and bio-error, and the quote of his that resonates with me is this: "The main thing that stands between the human species and the creation of a super virus, is a sense of responsibility among the individual biologists (11:32 - 11:41)."

The danger is that we'll become vulnerable to the ethics of ever-larger groups of people. We were vulnerable to the ethics of two people in 2011. They were good guys – they were actually virologist. Today, there's probably hundreds of people who – enabled by CRISPR – could do the things that were done in 2011. And at some point, that's going to extend to the world's elite life science grad students. And at some point, it's going to extend beyond them.

**Naval (12:05):** It seems like this is one of those fears that is almost built into the human species. We all realize this at some level. Everyone watches zombie movies, why? It's basically warning you what to do when the plague comes. Now, when the

real plague comes, unfortunately I don't think people will look grey and green and be lumbering around slowly. They'll be your loved ones, coughing and feeling sick very quickly. Or, if it's an engineered virus, it could be something that shows no symptoms until everybody drops dead 10 days later.

We do have some sense of alarm already built into our science fiction genres. I would argue that science fiction and movies actually warn people. Peter Thiel famously critiqued sci-fi because, he said, "There's no more optimistic sci-fi," in the '50s and '60s we had all this optimistic sci-fi. Star Trek was very positive and optimistic but nowadays it's all apocalyptic and negative. And I think that's because a lot of film makers rightly recognize that we are approaching a singularity but a singularity of a different kind, where a single individual can end everything.

**Rob (13:01):** And particularly when we start talking about synthetic biology, I worry more about the lone nihilist than the organized terror groups. Because with relatively few exceptions – and really the only exception I've been able to identify was called Aum Shinrikyo. They famously attacked the Tokyo subway with sarin gas back in the '90s. But generally speaking, terror groups are discriminate by nature and the kind of attack that you would carry out with synthetic biology or perhaps a nanotechnology some other things are indiscriminate by nature. So I do worry more about the lone nihilist.

**Naval (13:36):** I'll give you another scenario, I don't think it requires a lone nihilist. I think you can actually easily assemble a group of nihilists.

**Rob:** Once again, you've said, "pessimistic as you are, Rob, you're being too optimistic here." So tell me how you assemble –

**Naval:** I'm not trying to be a pessimist, I don't want to create world war three here.

**Rob:** No, the reason why we're in this conversation is I don't think you're a pessimist, I actually think you're a realist.

**Naval:** Yes. I'm a big believer in human freedoms and I'm not convinced that we need to start forbidding things.

**Rob:** Quite the opposite, as we'll discuss.

**Naval (14:00):** Right. But, I think it's quite easy to organize a group to work on this. There are multiple groups in history that have been wiped out, that have been eradicated, genocided. And very often, groups of people have essentially all committed mass suicide because they knew they were going to die. There's a famous one with some Jews who were surrounded by the Romans and rather than let themselves be enslaved by the Romans they killed their own families, drew straws, and then killed themselves. When India was being invaded by Islam, castles that were under siege, when they were falling, everybody inside would kill themselves.

Now the worst scenario – again, popular movies and literatures are so far ahead of us – is the Dr. Strangelove scenario. *Dr. Strangelove* the movie came out after the nuclear weapons were created in world war two. And the fear was that the US and

the Soviet Union had created doomsday devices. Doomsday devices basically said, "Don't try and kill me because if you kill me this automatically goes off and kills everybody." The problem is that doomsday devices will not stop at the nation-state level, they'll propagate to the group level (14:53 - 14:57). So if you are in a subgroup, a minority, that is being eliminated and eradicated – what you may do is create a doomsday device. And you say, "If we get taken out, everybody gets taken out." This is where you can get a group of intelligent people together to work on the problem.

**Rob (15:12):** That's really interesting.

**Naval:** So you can have a proliferation of multiple doomsday devices.

**Rob:** That's very interesting. Using the Masada example, which is the one for Jewish history, you could hypothetically say if the state of Israel was about to get annihilated--

**Naval:** Exactly, they might unveil a doomsday device. They might say, "Oh, By the way – we've got one of these." Then it turns out so do the North Koreans and so do the Kurds and so does ISIS and so on and so forth.

**Rob:** Then so do smaller and smaller groups, it might be the Rohingya in Burma, in Myanmar, who decide to create such a thing.

**Naval (15:38):** So if you can get sufficient group loyalty, it's very easy to create a suicide cult. Even the anti-natalists – they're extreme environmentalists who might go in on this because they're like, "Well humans are kind of the plague that's infesting the earth, let's get rid of humans and it goes back to paradise." And again, movies warn about this. If you remember *12 Monkeys* - Terry Gilliam's famous thriller about a synbio attack. Or *28 Days Later*, the zombie movie.

**Rob:** And one reason why my delusion that it would be really be nihilist individuals that we needed to be concerned about was soothing to me, is because groups are, of course, far more dangerous than individuals. Because they can pool skills and they can be in more places than one at once.

**Naval:** They can do it with current, or close to current technology.

**Rob (16:19):** Yes. And when I think about the lone nihilist I worry about people like the person who shot up that theatre in Aurora, Colorado. It was a *Batman* premiere, the guy who famously dressed as the Joker. He was an elite life-sciences PhD student. So he was in that relatively narrow circle of people that we would have to worry about. Because as time passes and a group of people grows the odds of somebody from that group doing something terrible, and bizarre, and completely off the rails goes somewhere from being slim, to being somewhere like high, to being somewhere like near certainty. So, we've posed this terrible instability that we believe, to varying degrees, lies in the intermediate future. You're probably more pessimistic than I am. I believe there are things that we can do to preclude this.

**Naval (17:02):** Solutions. I have a set of potential solutions but it's just not the world I want to live in (17:04 - 17:28). So I'm in denial in that in the sense.

**Rob:** I don't want the extreme ones. And so, here's what I think we need to do. You talked about Dr. Strangelove. I think that's a very powerful example. We did survive the Cold War, and I think part of the reason for that was, we got very good at telling ourselves ghost stories. We got very good at freaking out in a way that turned out to be very, very productive (17:16 - 17:29).

**Naval (17:29):** That is a great point. I mean, the first thing I worried about, even doing this podcast, is: should we be sounding an alarm? Do we want to be alarmist? The problem is, if you stick your head in the sand, it's not going away. As you said, this is on the Carlson curve?

**Rob:** The Carlson curve, yes.

**Naval:** Which is much steeper than Moore's Law curve?

**Rob:** Much steeper.

**Naval:** This is not even like AGI, where it's theoretical. It's inevitable.

**Rob:** It's coming.

**Naval (17:49):** So at this point, one way out of this is to create such an allergic and immune response in every human that when they get the idea that someone is capable of something like this, that it just creates an immune response for the whole species, right? Where essentially, every cell in the body becomes a white blood cell. And that may buy us quite a bit of time. I have young kids. I want people to be aware. I want people to be on their guard for it, rather than having some horribly, catastrophic thing that happens, that's sort of a near miss, which wakes us up. Because the crack down and the wake of that could be far worse. We lost 3,000 people on 9/11 and we're still paying for it today, in terms of the lost freedoms and all these extra wars that we ended up fighting. Better to be informed and prepared.

**Rob (18:33):** You've just said two things that make me think of immune systems. First of all, you had said all of us could be white blood cells to counteract this kind of attack. And I really love that analogy because first of all, obviously, none of us would survive even a day without our immune systems. And our immune systems, they got trained up when they get exposed to new viruses – and training the societal immune system, to start having some awareness and some concern about these scenarios now, is maybe akin to giving a vaccine of some sort.

**Naval:** Yes. It's a very interesting problem and immune systems have to be trained very, very carefully.

**Rob (19:10):** Then the other thing is what you just said about 9/11. That's the immune system overreacting. If you die of a sepsis infection, it's usually because your immune system overreacted to an extent that it started causing damage to your organs.

On a more narrow level, there is something called a cytokine response storm. That's something that happens sometimes in immunotherapy. Cytokines are things that signal the immune system to more and more white blood cells, and sometimes you

get into a situation where there is so many of these signals going out to make more and more white blood cells, that the body essentially creates a crippling number of white blood cells, and then kills itself.

And so, the 9/11 example you just cited is an extreme case of that. We didn't nip the original infection in the bud, and then we had this radical sepsis-like overreaction, spending trillion of dollars and killing a lot more than 3,000 people.

**Naval (20:17):** The Patriot Act, what it did to our freedoms, which are still going on to this day, and which just live now in a surveillance state, which is now all forgiven because of 9/11. In fact, when 9/11 happened – that day at work, I said to one of my co-workers, "There's gonna be a war over this." I didn't know who. I didn't know where. And I was wrong. There were *two*. In a drone strike on a commercial aviation scenario, you could easily imagine that the overreaction that we have afterwards may be worse than the actual, initial problem.

**Rob:** That's why you have to follow this very, very careful and narrow balance between training the immune system to recognize a new threat. Which is really all what we are talking about today – a lone wolf or a small organization doing terrible things with new technologies. But the reason that you really have to train that immune system is if even a failed attack gets through, the overreaction could be what really, really cripples us. If you consider the worst possible synbio attack to be something that is eradicating of all us. Something that could be a trivial failure on that scale – It kills 1% of us – could just cripple society.

**Naval (21:05):** It's very hard to do these counterfactuals. For example, 9/11. The response to that may be an overreaction – but, in a parallel universe, it could've turned out if we hadn't had that reaction, then they would've gotten their hands in a nuclear weapon. It's impossible to say, so the immune system is going to react the way it's going to react. But I agree with you in that training it is very important. We don't want everyone to be Big Brother. Every neighbor spying on everyone and constantly surveilling and reporting on everyone. That's a nightmare scenario. But at the same time, these new categories of technologies are just enabling new categories of threats that the immune system has to be trained to recognize and then respond to.

So what are the things that we can do? I like your idea of ghost stories. For example, *Ex Machina*, if you remember that movie.

**Rob (21:48):** Brilliant movie.

**Naval:** Brilliant movie. It's about an AGI threat. We have the zombie movies. It's been a while since I've seen a good synbio threat movie.

**Rob:** It has been a while, so to just go briefly back to the Cold War example. We did tell ourselves this terrible ghost stories with names like *War Games* and *Dr. Strangelove* and *Failsafe* and *On The Beach*. And as a result of that, everybody knew what the worst case stakes were.

**Naval:** There was a time shortly after World War II when both the Soviet Union and the US believed that a nuclear war was winnable.

**Rob:** Exactly, and we cured ourselves of that by talking about it and fearing it and working very, very hard on contingency plans.

**Naval (22:26):** It took scenarios like nuclear winters to show that even a small number of nuclear weapons being detonated in the atmosphere could end all life on earth. When it comes to plague stuff, we've lost track. Plagues in human history used to be far, far, far worse. There was a plague in 536 – they have now discovered 536 AD. Between a third and a half of the people in the Eastern Roman Empire, the Byzantine Roman Empire, were believed to have died from plague. Think of a half or a third of all the people you know dying of plague.

**Rob:** What's chilling about that is just thinking about how less resilient society is today. If you can imagine some kind of a synbio or other attack grievously sickening, let's say 1% of the population, most cities don't even have a fifth that much hospital bed capacity. Whereas in the Middle Ages, people were tough enough to shake off the deaths of five, 10, 20, 30% of their neighbors. I don't know what happens in an American urban area when 80% of the people who are stampeding a hospital, life threatening situations, are turned away for lack of beds.

**Naval (23:32):** One analogy this World War II. World War II was highly devastating for communities that were fairly built up from a civilizational perspective, not as much as we are today. They seemed to toughen up when they needed to because there is no other option. But, we are now much more densely packed in urban areas. We have much more complex supply chains for even basic necessities like food and water and power. So if we were out of food and water and power, for example in Manhattan, it would be a complete disaster. Much more so than it would have been in the Manhattan of 100 years ago.

**Rob:** Yes, and it seems like we have these single service dependencies, for lack of a better word, today. If the power goes out – completely out – that shuts down pretty much everything. The food starts spoiling, the cops can't communicate with one another, there's no way to navigate after dark. All exits from a city are going to be blocked.

**Naval:** We use civilization to put humans in more precarious locations than they would have been otherwise. Whether it's living on the floodplain, or living in the desert, or living behind bridges, behind single roads, or air delivery – there's just a lot more humans who could be stranded.

**Rob (24:37):** Also, thinking about what it'd be like to be invaded in 1939. I would imagine even in Northern Europe, there were very, very large sections of cities that were unelectrified or, at least, people that were living without electricity, probably had fairly recent memories of being without it. People were just much closer to living more directly off of the earth and more directly by their hands and the sweat of their labor than we are today. A biological attack that fell radically short of the worst case scenario could just be so devastating because of the knock-on effect.

**Naval:** Well, biological attacks create a level of fear that it is ongoing that other kinds of problems don't. For example, if there is an earthquake, after the earthquake has passed, everybody comes out to help everybody else. If there's a plague or a



biological attack, everyone goes into hiding and nobody wants to help anybody else because they get sick, too.

**Rob:** Which brings us to this amazing term you came up with when we were exchanging emails about my series of essays on these topics, and later tweeted, and it resonated with a lot of people which was the 'Ender'.

**Naval (25:38):** Yes, I believe that if you want to get a message out, you have to brand it and you have to put it in terms of people can easily synthesize, understand immediately and transmit. Like you said, the *Terminator* scenario, the Skynet scenario for AI – everybody knows what that is, so it's very easy to communicate. To me, the idea of an individual trying to kill everyone on earth – this person is an “Ender.” I'm riffing a little bit off of that book *Ender's Game*, it doesn't quite mean the same thing.

**Rob (26:03):** I thought it was a brilliant term when you came up with that. I'd been thinking about these topics – I won't say “nonstop.” But an awful lot for a couple of months, because I was writing this for long essays, for Medium, that wrestled with it. And I'd sent you the first draft of the first one or something like that. And we started lobbing emails back and forth. And you came up with this really really powerful term because it is so succinct. It has an echo of something that we're familiar with.

But we've just spent most of the last hour defining the Ender. And so, it is a complicated topic that could benefit a great deal from traveling with just a couple of words and a few syllables. Another analogy, I think I've mentioned to you before is this phrase, 'the tipping point'. It's a powerful phrase. Probably most educated people understand what it means and have used it. And the fact is, that term has a birth date. It was coined by Malcolm Gladwell in an essay in the *New Yorker*, probably 20-something years ago. He defined what a tipping point was. He came up with the term, and it's a great little essay, later titled a book after it.

I'd be surprised if more than 1% of the people who successfully use that term actually read the essay. And it's a complicated term and it conveys a great deal. I can now say, "Hey, you know, the tipping point in this market, or in this political situation, or whatever was, came at such and such a time when such and such happened." Now, all of the sudden, thousands of words of defining something has been collapsed into something very concise, where sophisticated conversations can be held. The trouble is, does it become aspirational? Like, is the term too good?

**Naval (27:31):** Right, we don't want it to become aspirational. It's not something I want to see exist. Nor do I want to turn it into some sort of biblical prophecy, where then, people are looking forward to the Ender. Because there's always some psychopaths out there who are. But I think it's important for us to recognize that these kinds of people exist. I wouldn't lionize it. It's not a positive thing. This is obviously a terrible, unloved, unlovable person to the extent that if they succeed, obviously, no one's around to-

**Rob:** Applaud them.

**Naval:** -applaud them. And if they fail, they'll be the most hated, reviled person in history. They're not doing themselves or their progeny, or their friends or family, or

even their ideologies any favors, because they're forever going to be associated with a horrible thing. And let's face it – most of them are going to be inept.

**Rob:** It's one more chance to fail.

**Naval (28:15):** So I think every time any one of them fails, the idea would be to not give them any attention. But as we've shown through terrorism, we don't really do that. We give it a lot of attention. But I think the way to look upon this character is like a Gollum-type character from *Lord of the Rings*. This pathetic, evil, bumbling, inept character that no one should aspire to be. But at the same time, it is important to get the message out. We need terminology and vocabulary to be able to communicate. We need to be able to transmit the information, to absorb it, and to keep an eye out for it. And if someone says, "Hey, I think so and so wants to be an Ender", that gets the point across to how we should investigate that person.

**Rob (28:49):** Gollum is actually a really, really good example. Nobody wants to be Gollum when kids go around playing *The Hobbit*.

**Naval:** Yes. The real Ender is likely to be a Gollum-type character. Someone who is warped and distorted, unhappy and miserable.

**Rob:** Another term might be, for the failed Ender – I was thinking last night – enderlings. [laughs] An enderling is somebody who tries and fails. And most people are going to fail spectacularly in this objective.

**Naval:** Some combination of a credible attack by an Ender –

**Rob:** *Enderling*.

**Naval:** Thank you – enderling. Combined with the cytokine response could take out a lot of humanity and civil society as we know it.

**Rob:** Which is why we need to marshal our immune system against enderlings, because there's going to be thousands of them for every Ender. And hopefully, there'll be an infinite ratio, because – I think, if we think hard enough about it and early enough about it, there won't be an Ender. But people have been attempting to kill as many people as possible on the act of dying for as long as we have records.

**Naval (29:48):** This is going to be a little sappy but my conclusion is that you're not going to be able to stop them purely through surveillance, or creating a totalitarian society, or Big Brother watching, because that creates its own immune response. And you're not going to be able to stop them by stopping technology, for all the reasons we've already talked about. So, really, the only way is to not create them in the first place. And to not create them, means that we all just need to learn how to make sure nobody goes unloved. The one part where the immune system analogy breaks down a little bit here is we're not talking about an external threat to the human body. We're talking about an internal threat to the human body so it's more like a cancer cell.

**Rob (30:11):** Cancer cell, yes.

**Naval:** Exactly. So don't let cancer arrive in the first place.

**Rob:** Although one of the things that I learned at this wedding over the weekend is that our immune system actually is taking out cancer constantly.

**Naval:** That's true.

**Rob:** It does, successfully, rally itself. And it's only when cancer gets out of control and figures out a way to foil the immune system that it goes away. So a cancer cell is still something that an immune system can rally against. And I think, you've really just hit on the first layer – if we can talk about it like a communication stack. The first layer is making sure that people don't feel unloved and alone.

**Naval:** Prevention in the first place.

**Rob (30:54):** That's the first layer. The second thing is also really inexpensive. I'll just call it "imagination," for lack of a better term. The 9/11 example is a really, really strong one where we lacked that least expensive input of imagination. When I look at that, I can't say that it was a lack of foresight because we had actually hardened the airports throughout the world in reaction to terrorism in the 1970s. It wasn't a lack of resources because the world governments have probably literally spent trillions of dollars preventing a sequel to 9/11 since it happened, so we can't say we were impoverished for resources.

We can't say it was for a lack of brain power because although Osama bin Laden was very charismatic, he had a great deal of money, he had a lot of strengths; nobody ever thought that he was an Einstein-level genius. So he came up with an *non*-ingenious plan. And the team that he then assigned it to is fairly under-powered. They're not very comfortable navigating in this alien society that they find themselves in, probably not any more than I would be, trying to navigate my way through Saudi Arabia. They're a fairly scrawny lot. They're not all that bright. A lot of the people didn't even realize that they were on a suicide mission. And they pull this thing off.

On our side, we had the resources, we had the foresight, we had the brainpower. But we never give anyone the job of supplying the imagination. We might have – I don't know. Populated a conference room with a bunch of really smart people from a bunch of different cultural backgrounds, probably over-representing parts of the world that it had a lot of experience with terrorism. In 2001, that would have been Sri Lanka, maybe still Ireland, certainly the Middle East.

You know, as a country of immigrants, we get to tap into all kinds of perspectives – and just said, "Hey, you guys in this conference room. What you do 40 hours a week, 60 hours a week – we can give you all the pizza and beer that you need. You just sit down and think *really* hard about the low-tech, cheap things that relatively underfunded, understaffed, under-geniused organizations, like Al Qaeda, might be able to pull off. And just use your imagination."

Like, "OK, We're dealing with terrorists. What have they done a lot of?" "Well, since the '80s, there've been a lot of suicide attacks, starting in Lebanon kind of in a military situation." "OK, Suicide attacks. Well also, they hijack a lot of planes." That's thing number two. Maybe you have seven things on the board. But you don't have

1000 things. Then you start thinking, "Well, what's some crazy suicide airplane thing? Have we ever seen such a thing?" And then hopefully somebody would say, "Yes there were 3,000 of them in the biggest war in American imagination, World War II. They were called Kamikaze." That doesn't feel like a terribly expensive investment. Just having some smart people ideating on, what could be done with the tools that we have now and the tools that we have five years hence.

**Naval (33:40):** Generally, governments only tend to react to something after it's happened, they're not proactive. They're not known for that. Private industry tends to be better at being proactive. This Gatwick drone incident does increase the odds that now you will have governments thinking about what do drones mean for commercial aviation. I should add as a disclaimer that I am an investor in one company, SkySafe, that helps with security against drones, but it's a tiny investment. The reality is, although I think they're good for some things, I don't think they could stop many of the scenarios that we mentioned from happening. And there are other great drone security companies out there as well. I think DroneShield is another one of them. I think AirWare has some software.

**Rob (34:16):** Yes, and correct me if I'm wrong, you probably have several hundred small investments in various startups.

**Naval:** Yes.

**Rob:** Something like that.

**Naval:** That's correct.

**Rob:** So I'm personally confident that there is zero chance that this entire conversation has been an ingenious manipulation by you to get SkySafe equity more valuable, but you are very--

**Naval:** If that was the case, they owe me a lot more stock.

**Rob:** A lot more stock. But you are very, very good to note that. It's always good to hoist the flag over potential conflicts.

**Naval:** Back to my original point, the drone incident at Gatwick means that governments will start trying to develop an immune response.

**Rob (34:45):** Yes, governments might start getting ham-fisted and reactive. But we can also probably count on companies like DJI, really powerful drone companies, to say, "Oh my God, I don't want to be eradicated here. I don't to be regulated out of existence. I'm going to start an immune response as well." An interesting example might be the film industry. There was concern that the government was going to start regulating. So the industry came up with its own PG, PG-13, rated G, et cetera standards.

I think that the best defense, particularly when we talk about synthetic biology, can and should come from the industry itself and even from the professors in universities that train the people who are going into the industry. That euphemistic conference room, in which people are constantly thinking about what could be done and what we

could do now to protect against this – that should be part of the discipline of synthetic biology. And if that is, to some degree, formalized in training – and it's formalized in industry associations, and maybe something like the W3C, only for synbio, that has regular meetings. That could be, potentially, our most powerful weapon against this.

**Naval (35:49):** Something like AI, less likely much further away, but it's scary because there's no time to develop an immune response.

**Rob:** It's too late once it's happened.

**Naval:** It's instant when it happens, exactly. Even something like a major synbio attack could get to a catastrophic level very quickly. We haven't really seen minor designer synbio attacks yet.

**Rob (36:09):** Well, I'll actually give you the example that kind of proves the rule. The anthrax attacks in 2001 are really sealed in my memory because as it happened, I was in the Senate Majority Leader's office, Tom Daschle, the day that a letter arrived at Daschle's office with anthrax in it. That particular letter didn't infect anybody, but several people were killed. And it ultimately was traced back to a very, very senior weapons researcher. That's a chilling precedent, when you think about it. We were probably as well-organized of a military industrial complex as the world has ever seen.

And we – even with all of our good intentions and open society and strong protection measures – we couldn't prevent some highly weaponized anthrax from getting out of one of *our* bio labs and into the office of the senate majority leader! *We* couldn't do that. So how can we expect an entire diffused industry of tens of thousands, hundreds of thousands of synbio experts – how can we expect all that diffusion to keep a lid on itself? It only can happen with a multi-level immune response, starting with looking after people so they don't feel unloved, secondly by infusing this imagination into the industry and also into government and elsewhere.

**Naval (37:19):** Then next comes the training of the immune system. Which is, any attack that does happen or was attempted to happen, we figure out what that looked like, what we can learn from it, and we start training pockets of people around the world to keep an eye out for these things and respond to them. Ideally, we could also crowdsource it; filtering the air for any unknown new pathogens. You had George Church, you had a great podcast with him, where he talked about essentially scanners that can figure out what's in the room. That can map viruses crossing the globe in real-time. And these can be privatized almost – like, buildings can install them. The new Teslas have these amazing bio-defense HEPA filters.

**Rob:** Really?

**Naval (37:57):** Yes, there's a Bioweapons Defense Mode in the new Tesla where you hit a button and the AQI comes down to 7 – even when I'm driving through the current smoke laden Bay Area. Why aren't those installed in every new house? These are all short to medium term, and I think these are all things that we need to do. But if you're talking about long-term solutions – this is a horrible dystopian one, I don't want to live in it. But you're born, and there is a chip implanted inside of you,

and it basically monitors your mood. And the moment you get actually suicidal, it just turns you off.

**Rob:** Talking about the storytellers who've prefigured certain things – that almost sounds like the V-chip of South Park, times 100.

**Naval (38:32):** Right, exactly. The problem is, this would be a totalitarian nightmare. But maybe there is a more benign version of it, where instead of surveillance from above, it's sous-veillance from below – where it's all of us keeping tabs on each other. Now unfortunately, this turns into a neighborhood watch creepy thing where it's like, your neighbors are watching you. But if someone is going suicidal – especially if they're a life sciences PhD student – you don't necessarily treat them as someone who should just get help. It's someone who should actually be watched and maybe even put inside an institution until they are better. I think there is the obvious practical stuff of just being far more careful and regulated about who gets access to these facilities, these technologies, and this know-how. But even that's only going to work for so long.

**Rob (39:17):** For me, I get my relative optimism, first of all from some of those examples from the past like getting through the Cold War. There's also the positive aspect that we outnumber this hypothetical person or group, that might want to bring the curtains down on humanity, by a factor of several billion to one. There's more of us, and we're thinking about this earlier. These people who detonate in a rampage murder way, when they go off, they tend to grab the implements at hand. That's why there's so many mass shootings in the United States, mass stabbings in China. Occasionally the thing that's at hand might be the throttle of an airplane, but this isn't an aspirational career.

**Naval:** It's true. You don't learn to fly a plane so you can fly it into the ground.

**Rob (39:59):** Exactly. When somebody snaps, as they have done in thousands of well-documented cases, they tend to grab the implements at hand with the training they already have. Now, sometimes, they might plan and scheme for a period of days, maybe even weeks, I don't know, maybe even months. But it's not like you have an adult Stewie Griffin – you know, the evil baby in *Family Guy* – that spend decades positioning him or herself to annihilate the entire world.

If Andreas Lubitz, the Germanwings pilot, wasn't a pilot, he might've gone crazy with a knife. Because he was in Germany, probably he couldn't get a machine gun at retail, right? He wouldn't have gone to flight school for four years and worked his way into the position that he could do his thing. So the immune system that rises up is going to be inside that environment where that person might be empowered. Again, back to synbio – nobody, I don't think, is going to become synthetic biologist so that after they get their PhD, and do their postdoc, and their professorship, and they get great access to a lab that'll have 20-years-from-now technology, they'll be able to end the world. No.

What we have to worry about is somebody in that inner sanctum snapping – and then doing something within a span of probably a few days, to weeks, to months, that is diabolical and catastrophic. And a much narrower set of things that could

potentially be done because of the inherently narrow timeframe. I mean, 99.99% something high percent of synthetic biologists sure as hell don't want this to happen.

They will come up with, and be able to access ideas that you and I can't because they are expert in their technology. One relatively simple thing that's already being done is if somebody is requesting a print out of the pathogenic sequence of DNA, the system will large should, A, refuse to make that pathogen and B, alert the authorities. Now, in this early time that we're living in right now, there aren't desktop printers that are capable of making full-blown viruses. But when an organization like Twist Bioscience, which recently went public – they work as a service bureau – and they will provide people with sequences.

When a pathogenic sequence is ordered – I don't know if one ever has been, but my understanding is that the protocol is if one is ever ordered – they ain't going to fill that order. And they're going to help people get to the bottom of what's going on. I think we can rely on the good intentions and the imagination of the crushing majority of people in any given field, like synthetic biology, to think through those catastrophic scenarios – provided that there is an institutionalized desire to use our imagination in that manner.

There's an institutionalized desire to think about the most twisted things people could do with the tools that are coming online before somebody else thinks of those things. That's why I'm confident that if this imagination – and it's a painful imagination, it won't be fun to think these things through – but if that's part of the discipline of synbio, most of the really crazy things will probably be thought of years in advance. And countermeasures can be taken that aren't that intrusive, but are just clever and foresighted.

**Naval (42:58):** There are a couple of ways that are far fetched, that I can imagine, even in the long term. One is we just invent some brand new technology that is defensive in nature. We did that recently with encryption. Encryption is much better for the defender than the attacker. The attacker has to brute force a nearly infinite number of solutions, whereas the defender can very easily protect themselves on digital encryption.

**Rob:** That's a really good example. So that is an asymmetry in which the defender is asymmetrically empowered against an attack.

**Naval:** Exactly. So the question is, can we invent something like that in synthetic biology? Is there an advance to be made, especially with a lot of resources, and a lot of good people looking at it, that could aid the defender?

**Rob:** And let's talk about what not to do. I think that the top of your list and mine is a technology ban.

**Naval:** Well, it's impossible. The problem is, if you ban technology, then a few countries would say, "We're still going to keep developing technology." Let's say North Korea and China go for it. Then all you've done is left yourself behind. So the only way to have a technology ban is basically to start World War III. [laughs]

**Rob (43:56):** Yes. And internally within our own society, it's worth noting that illicit drug labs are both illegal and ubiquitous. This is not something like the creation of a nuclear bomb, which needs to marshal the entire resources of a nation state. You could be doing amazing synthetic biology in your own living room, and the neighbors wouldn't necessarily notice.

The other reason that I'm personally very much opposed to a technology ban, is because I look at some of the things that are going on right now in synthetic biology, and I find it to be purely exhilarating. I believe that synthetic biology is on the cusp of curing the organ shortage. That's going to relieve so much death and misery from all quarters of this world. It's amazing. I think synthetic biology is going to make clean meat, which is going to be an ethical triumph for conscious systems in general, and also for issues like global warming. Synbio is going to accomplish so much. And throttling it in its crib would be, A, impossible, as you just pointed out, and B, a terrible immorality. It's not something we want to do.

**Naval (44:57):** Technology is a coin, and one side of it is immortality, and the other side is annihilation (44:57 - 45:01).

**Rob:** That's a very good way of putting it. Did you just come up with that?

**Naval:** Yes. [laughs]

**Rob:** I like that.

**Naval:** I'm a happy libertarian, but I think this is one of those cases where it does make sense for the government to pay attention. It goes back to your original point, which is, now the gains are privatized, the losses are socialized, and you're dealing with very catastrophic outcomes. Humanity is a public good.

**Rob (45:25):** The survival of humanity is the ultimate public good (45:25 - 45:29).

**Naval:** There are other good scenarios, if we want a little bit of hope. [laughs]

**Rob:** It's time for a little bit of hope.

**Naval (45:34):** Hope is in short supply. If you look at how we connect with our phones all the time, we're much more in-tune with other people and their feelings and where they are in their lives and so on. It's not too much of a stretch to imagine that our children who will actually get cybernetic implants, which will keep them connected to the internet at all times. Once you're cybernetically implanted and connected to the rest of humanity, you become almost like a cell in a multi-celled organism.

Now, that's voluntary. It's not some involuntary horrible thing. But now, your friends and family and loved ones can see your mood and see when you're about to detonate, for example. I think that there are support systems that we can create through technology that will help raise our consciousness to a new level, and our connectedness to a new level.



At some point, if somebody tries to end the world, it'll be because they weren't loved enough.

**Rob (46:17):** Yes. That gets to your other point of, let's not make more of these people who get to this horrible, horrible point. Easy as it would be to have feelings of hatred for Andreas Lubitz, the pilot for Germanwings, this was a person who was in horrible agony and pain and suffering – and did something reprehensible. We should not create people like that. If we're a gentler society, or if social media behaves in a more supportive and less tearing-down way – fewer and fewer people are going to be pushed in that direction.

**Naval (46:48):** Before, we used to find love and longing in our family and our tribe. Now we live in cities, disconnected from our ancestors and our immediate family. And we're looking for ways to connect. We're looking to create new tribes and new families, and I think people are using all tools at their disposal to do that. Burning man is kind of this crazy festival in the desert where people go and connect with 70,000 strangers as if they're family. They're radically inclusive. And a society like that is just going to have a lot fewer people detonating – and if they do they can be easily intercepted and noticed and cared for.

Generally, when someone is depressed and they kill themselves, deeply, internally they feel alone. We are headed towards a society of more and more connectedness. Whether it's through meditation, whether it's through spirituality, whether it's through psychedelics, whether it's through consciousness festivals, whether it's through parties, whether it's through social media whether it's through the internet, whether it's through cellphones. It doesn't matter. But the more connected we can be as humans, the better off we are.

**Rob (47:47):** I think there's also less reluctance to use certain tools that might be very powerful in counteracting these things. When we think in terms of these thousand people who detonate annually – even if society or the government takes the most self-centered approach to that, the inevitable conclusion is the things like treatment-resistant depression and treatment-resistant post-traumatic stress are societal dangers.

They don't just present the danger to the person who is suffering from PTSD or suffering from treatment-resistant depression. But they could present a danger to anybody in the movie theater if they go nuts with a machine gun or perhaps anybody in the world if they go crazy with a DNA synthesizer in 30 years. And it seems that they are some pretty powerful tools to fight these things.

There are advanced phase free trials in Europe using psilocybin for treatment-resistant depression showing an enormous amount of promise – and I interviewed the people behind that experiment in an earlier episode of this podcast. And simultaneously, in the United States, there is a phase three trial going on under the auspices of the FDA testing the potential for MDMA – molly or ecstasy – in treating treatment-resistant PTSD. We got this societal aversion to using these chemicals in any kind of a clinical setting in the wake of the 60s because of the horror that the psychedelic era has imposed on certain elements in society.

**Naval (49:09):** It's insane that we take any substance that's bioactive, when you take it in your mouth and it causes a change in your feelings, and we ban it. The reality is some of these make you much better off. So if you have a chance of detonating I would rather society spirits you off into a beautiful farm somewhere to hang out in the fields and do all the psilocybin and MDMA that you need until you basically get through accelerated therapy. That's worth it.

**Rob (49:35):** MDMA actually dates back to World War I. That's when the molecule was first created. And psilocybin, I think, was discovered by non-indigenous cultures around the 50s. LSD was first synthesized in the 1930s. These are very much in their 0.9 form. And then, we banned any kind of clinical research into their potential – and not only their potential but into their modification and their enhancement.

Imagine if all the muscle and budget and imaginative resources that the pharmaceutical industry has available to it – if they had spent 30 or 40 years trying to create the 2.0, 3.0, 4.0 versions of these molecules that we stumbled into back during World War I. Given that these seem to have tremendous potential against PTSD, depression and other things we would probably have far more effective remedies today. Well, it seems that kind of development and invention and imagination is about to be applied to these fields.

Maybe that is something that can radically reduce the number of people who are inclined to detonate.

**Naval (50:37):** I think related to that, meditation is making a huge surge in the Western world. And actually, the reason I was slightly late when you came today was because my meditation was too good. So I ignored the timer when it went off. And I kept going. That's not to humble or brag or anything, it's just that I've found that if you can really get into it, it's so good that it just brings a sense of peace and joy into your life that no one can take from you. And it's a shame that it's not really on offer like, let's say in high school.

**Rob:** Yes. Why not just be part of gym class for ten weeks.

**Naval (51:049):** Yeah, it's exercising your mind, but we don't teach that skill. It doesn't cost anything. You don't need anything. And it's probably better than most drugs over a long period of time and it's better than most therapies, I would even argue, at least from my personal experience. I think these kinds of consciousness raising tactics combined with prudent scenario planning; restricting access to some things that individual just should not have access to; creating a shared vocabulary; popularizing the concepts so that essentially every human has their own immune response to this and understands it.

Identifying when people are taking privatized gains and socializing the losses – calling that out and saying, "Hey, there's no amount of skin in the game you can have that makes it worthwhile, because you're putting our skin in the game without our permission."

**Rob (51:52):** Yes. Making that like the ultimate crime, Creating an enormous amount of social awareness and also social shame around socializing a cost.

**Naval:** I give credit to people like Elon Musk for starting to speak out on it and he's financed the OpenAI project and thinks he's going to get us to Mars. But that's not enough. He actually should even talk louder. He does talk. I really appreciate that.

**Rob:** And thank God for that. But, we need 10 more like him or a thousand more like him.

**Naval:** We need a thousand more like him.

**Rob:** So the antibodies that we can muster – first of all a gentler society that has fewer people detonating, fewer people killing themselves, fewer people being lonely. Is using the tools and allowing the pharmaceutical industry to develop the tools that are most promising in this area.

**Naval (52:30):** What if we rated the pharmaceuticals on how much they lowered suicide risks?

**Rob:** Yes. And maybe the fact that a very significant external risk is being posed by these types of mental states could raise the urgency around it.

**Naval:** I would also vote to have a Manhattan project around the kinds of scanners that George Church was talking about.

**Rob:** Bio scanners.

**Naval:** Exactly.

**Rob:** Yes. That give very, very early warnings leading.

**Naval:** We need to take it to the next level.

**Rob (52:51):** The other thing is, some people could listen to this talk and say, well, the last thing we want to do is have CRISPR education in high schools. I actually say the first thing we want to do is have CRISPR education in high schools. I think we need to acknowledge that the vast majority of people are good guys rather than bad guys. I've researched it pretty carefully, and I can think four instances in which commercial pilots have downed their planes intentionally killing everybody onboard. That's over a span of decades. There are tens of thousands of commercial flights every single day.

The crashing, overwhelming majority of commercial pilots obviously refrain from doing that, even though they could. The crashing majority of synthetic biologists would refrain from doing that if they could. And if part of their training was also thinking about what one bad seed might possibly do. If that was part of the discipline of synbio, then the more synthetic biologists we have the safer we are – not the more endangered we are.

**Naval (53:44):** Right.

**Rob:** We just need to have that imagination – slightly paranoid imagination, yes – but that imagination. Those ghost stories, these nightmares, be part of the training process. And that's how we create an immune system that can prevent this.

**Naval:** And this will pay benefits even if there is not an incident like the one that we're talking about.

**Rob:** Totally.

**Naval:** You'll just have less colds. You'll have less Ebola. You'll have less of every kind of virus and bacteria spreading.

**Rob:** I guess a big question is whether people and society, writ large, want to mentally engage with these dangers for decades. All the time we spent envisioning nuclear winter might have saved our lives, but it wasn't exactly fun.

**Naval (54:22):** I think just like with the tendency to not look at the screen in a horror movie and to not want to get bad news and to shoot the messenger for bad news. It's built in to us as humans and instinct to face away from it, to deny it.

**Rob:** Much as I enjoy hanging with you, I don't think this has been a very fun conversation.

**Naval (54:37):** No, this is not going to be a viral podcast, no pun intended. Because I think people don't want to spread bad news or even bad thoughts – because people think that if I don't think the bad thoughts it won't happen. But think of it sort of like as the union of concerned scientists came up after nuclear weapons were invented. These were people who knew them well. They knew what they were capable of and they did their best to get their message out. They built the doomsday clock. They worked tirelessly against nuclear proliferation.

They did come up with the horror scenarios of nuclear winter ending all life on the earth and so on. They had to popularize that and, yes, there were some over reactions. There were kids doing drills, running in the basement and hiding underneath their desks in case of a nuclear strike, but we needed that to get the message of nuclear nonproliferation out there very, very strongly. And I would argue that nuclear proliferation is a genie that has been kept in the bottle to a much greater degree than anyone imagined in 1950s.

**Rob (55:28):** Against all odds. What smart observer in 1946 would have imagined that 70 plus years would go by without a nuclear weapon being fired in anger?

**Naval:** Right. So it's better to have a slow, measured, trained immune system response than wait until something happens. Then there's a cytokine storm. Then there's multiple wars, surveillance, crack-downs, police state. That's not where we want to end up. We want to be a free and civil society. And so that means having a measured approach to this rather than being reactive to this.

**Rob (55:57):** The nuclear example is an incredibly powerful one and triples my confidence that hopefully getting this term Ender out there in a way that people understand the term and spread it, and hopefully people not shunning this odious,

grim conversation that we've had and spreading it to some degree can spread the kind of awareness that we need. One of the most amazing things in geopolitical history, in my opinion, was in 1990-'91 after the fall of the Soviet Union, four newly-born countries voluntarily denuclearized.

I think it was Ukraine, Kazakhstan maybe it was Uzbekistan and I think one of the other post-soviet republics had nukes on their land. How improbable, when you think about it, for four countries-- that's probably more countries that have nuclearized since 1998. I think there's really only been three. India, Pakistan and North Korea. And the reason that they did that – this crazy geopolitical move – was because humanity had spent 45 years telling itself ghost stories about what could happen and what we wanted to avoid. That's an amazing, nigh miraculous thing that did happen because we were brave enough to think about it and worry about it.

So these disaster-modeling exercises, which are kind of like prophylactic nightmares, can be very effective. Of course, let's hope the synbio-nightmares and the AI nightmares we're promoting right now will look like totally unnecessary exercises 50 years from now. Like the duck-and-cover drills that schools used to do during the Cold War. But we're happy when precautions turn out to be unnecessary. If we fasten the seatbelt when we start our car, we're not annoyed if we don't get into an accident.

**Naval (57:40):** There's a great Arabic aphorism that I love that says, "Trust in Allah, but tie your camel."

**Rob:** [laughs] That's good. Well, thank you so much for spending so much time and having such a crazy brainstorming conversation with me.

**Naval:** Thanks. On the one hand I hope we didn't terrify people, on the other hand I hope we did wake them up a little bit.

**Rob:** Yes, it's time for us to start thinking about this stuff.

**[01:56:17] [END OF AUDIO]**