FREE FORUM with TERRENCE McNALLY - A WORLD THAT JUST MIGHT WORK

LESLIE VALIANT recorded 03-31-2025 - Transcript

THE IMPORTANCE OF BEING EDUCABLE: A New Theory on Human Uniqueness.

McNally
Hello, I'm Terrence McNally. Welcome to Free Forum: A World That Just Might Work. I'll be speaking with Leslie Valiant, professor of Computer Science and Applied Mathematics at Harvard, recipient of the Turing Award, and author most recently of *The Importance of Being Educable: A New Theory of Human Uniqueness.*

On Free Forum, we explore the lives, the work, the ideas of individuals that I suspect have pieces of the puzzle of a world that just might work. We look at politics, economics, environment science, health, culture - all based on the fact that I believe we can do better and I want to find out how

The show streams weekly on the Progressive Voices Network on tunein.com. Podcasts are available anytime, anywhere on Apple Podcasts, Spotify, most podcast sites. And at my site, terrencemcnally.net.

I've been doing this for over 25 years and people often ask me, what are my favorite interviews? And of course that's pretty much impossible to answer. But I have found that among my favorites certainly are the ones I've done on science and with scientists.

I think one of the reasons is that it's the subject I know the least about. I'm pretty generally aware, knowledgeable, I can fake it with politics, even economics, culture to a pretty good extent. The kind of things that generally come up in conversation that you do a lot of interviews about. I can do that.

Science is a different story. In college at Harvard in the late sixties, they had a General Education requirement in Science. I had to take a science course. If this word makes sense, I looked for the “softest” science I could find. I didn't take a chemistry course, I didn't take a physics course, I didn't take a serious biology course. I took a course on evolution.

Now I'm really glad I did because I think looking at the world through that lens is very useful. And I confess, I'm not sure if a field like evolutionary psychology even existed at that point, but I believe there's much to be gained by looking to evolution to explain human behavior, human development.

So getting back to my experience of science-related interviews, I was hired almost a decade ago by Harvard's Wyss Institute of Biologically Inspired Engineering to host and produce with them a podcast series. We called it *Disruptive*. You can find it online. We did 15 episodes and we were honored by the Webbys In 2017, as one of the top five science podcasts, I engaged with innovating scientists on their cutting-edge work in DNA programming, robotics, novel treatments and therapies, and the like.

Today's conversation with computer scientist, Leslie Valiant, also explores a question I find myself returning to over and over again over the years, what makes us human? What unique abilities have allowed homo sapiens to succeed, flourish and dominate, knowing it's not our size, our strength, our speed. And one of my upcoming conversations actually will be with Vanessa Woods about the ideas in the book she wrote with Brian Hare, Survival of the Friendliest: Understanding Our Origins and Rediscovering Our Common Humanity. I'll tell you just a little bit about that one to set up this one.

Their conclusion - what made us evolutionarily fit was a remarkable kind of friendliness, a virtuosic ability to coordinate and communicate with others that allowed us to achieve all the cultural and technical marbles in human history.

And they acknowledged that this gift of friendliness came at a cost. Just as a mother bear is most dangerous around her cubs, we are at our most dangerous when someone we love is threatened by an outsider. The threatening outsider is demoted to subhuman - fair game for our worst instincts. And Hare's research reveals that the same traits that make us the most tolerant species on the planet also make us the cruelest.

Going back to the evolutionary perspective, let me comment briefly on the notion of survival of the fittest. I suspect that the popular understanding is that fittest means strongest, most physically dominant. It most certainly does not. What Darwin was talking about was fitness to survive, reproduce, and remain. And fit in those terms may have nothing to do with what we think of as physical fitness. Rather think of fit as when a puzzle piece fits in a puzzle. I think that's a way of thinking about fit and fitness that's much more in line with Darwin's thinking.

Leslie Valiant believes that if we hope to share our planet successfully with one another and with the AI systems we're creating, we must reflect on who we are, how we got here, and where we're heading. The new book lays out some steps along that path.

First, what he terms the educability of the human brain can be understood as an information processing ability that sets our species apart, enables the civilization we have, and gives us the power and potential to set our planet on a steady course. Yet, like Hare in Woods, he acknowledges that this gift comes hand in hand with an insidious weakness. While we can readily absorb entire systems of thought about worlds of experience beyond our own, we struggle to judge correctly what information we should trust.

Valiant argues that understanding the nature of our own educability is crucial to safeguarding our future. And that education should be humankind's central preoccupation so that we don't indulge our view of vulnerabilities by falling victim to propaganda and informational and emotional manipulation. Sounds kind of timely, doesn't it?

LESLIE VALIANT is the T. Jefferson Coolidge Professor of Computer Science and Applied Mathematics at Harvard University. Recipient of the Turing Award and the Nevanlinna Prize for his foundational contributions to machine learning and computer science. He is a Fellow of the Royal Society, a member of the National Academy of Sciences, and the author of *Probably Approximately Correct*; *Circuits of the Mind*; and his latest, THE IMPORTANCE OF BEING EDUCABLE: A NEW THEORY ON HUMAN UNIQUENESS

Welcome, Leslie Valiant, to Free Forum: A World That Just Might Work.

Valiant:

Well, thank you very much for inviting me,

McNally:

And let me tell listeners we're recording this conversation on Monday, March 31st.

Now, Leslie, I like listeners to get a feel for the people behind the work and the ideas. I'm lucky enough to have an hour here to play with. So can you tell us briefly in your own words about your path to the work you do today? And feel free to go way back. Sometimes people mention childhood inspirations, we're looking at that sort of thing, turning points, moments of decision. Take it away.

Valiant:

Yes. So since you started with your relationship to science, I can start with that too. I've always been interested in science since childhood, but all the science I did in college undergraduate was to do with the physical sciences, the physical world. And so it was a revelation to me that one could do something else as well.

I remember a year after my undergraduate degree, I was looking for something to do for the rest of my life, and I wanted to do something mathematical, and I discovered the work of Alan Turing. And he had - out of the blue - made statements about… mathematical statements not about the physical world, but about human capability. He proved that certain things are not computable.

So when I read this, I realized that this was really totally different from anything I'd seen before and kind of intrigued me. And I think this single episode which happened in the library made me turn towards computer science. Of course, much of computer science is really kind of about computers, the capabilities of computers. But after a while, I returned to this theme - that may be the most exciting thing about science is to push the boundaries, to have science applied to some area where previously people thought science just couldn't be applied…

McNally:

Right. In other words, the thing that you discovered yourself, you felt you could then run with.

Valiant:

Yes, yes. So then when I started working on learning, I think this was an area which at the time many, most people thought just was beyond science. So human behavior as far as learning was something which was beyond science, and I shouldn't be doing. But, of course, I and other people did, and it has had consequences.

So this is where, at the moment, I believe that mathematical sciences, particularly the computer science angle, has a lot to say about human behavior. If we want to understand ourselves, then traditionally we look to the humanities. But I think there's also another angle which will be increasingly fruitful, and that's kind the area I'm coming from.

McNally:

And if I'm not mistaken, you said computer science can help us understand ourselves, but you also very critically make the point that our making strides in computer science, AI, for instance, will actually call for us understanding ourselves better. It works both ways.

Valiant:

Exactly. Exactly. Actually, even Alan Turing said this, that one reason why making machines which are like us may be not a totally ugly thing to do, is because on the positive side, it'll help us understand ourselves.

McNally:

Very good. Yes. I wanted to…

Valiant:

I truly believe that.

McNally:

Just a little bit about Turing… You won the Turing Award. First, who was he? For people who don't know…

Valiant:

Well, he was a English mathematician. He was a pure mathematician who started early thinking about the capabilities of the human brain. And so he revolutionized science by writing this critical paper which founded computer science. And then during the Second World War, he also did fundamental, well, very important work in breaking codes.

McNally:

And there was the movie THE IMITATION GAME about that code breaking, which included Alan Turing…

Valiant:

Right, right. So besides being very impactful as a scientist, you're also very impactful through public service, if you like.

McNally:

Exactly. And you've sort of said it, but before Turing's time, computation was thought to be a psychological process done mostly by humans, but his formulation expanded that view.

Valiant:

Yes. So he defined it in the abstract. In some sense, he said there's only one kind of computation. Neurons and human brains do it, and maybe one can make machines to do it. Because at the time they weren't, machines weren't very powerful. And you'd find the notion which was independent of the hardware on which you realize it. So we had a notion, which we still agree, that if you prove that something is not computable, that it's not computable either in a machine or in biology.

McNally:

And when he's talking about hardware, we are one form of hardware.

Valiant:

Exactly. Exactly. Exactly.

McNally:

So what did you win the Turing Award for?

Valiant:

It was more than one thing, but the relevant thing here - it was for formulating what machine learning is. So what's called the "probably approximately correct" model, - which you quoted as a title of a book - is basically a specification of when you should regard a method of learning to be successful. So it's specifying what you want to get done, not so much how you do it, but what do you want to get done?

McNally:

Okay. You've been celebrated for your willingness to address some of the unsolved problems of science. And my guess is that the whole notion of solvable and unsolvable problems is new to many. I mean, they know it in their regular lives, but that this exists in science - what is that about?

Valiant:

Solvable and unsolvable problems?

McNally:

Yeah, I'm just saying, to a layman, that seems like, of course it sort of makes sense that there would be things, but I don't think we think about it the way scientists would think about it.

Valiant:

Yeah. Well, so in general, science, what's solvable and unsolvable, that's hard to specify. But in terms of mathematical problems… The surprise there is that - which is what Turing explained - is that just because you define a problem very precisely so that everyone understands what the problem is, that's not enough to guarantee that there's a solution.

Many problems you can't solve because it's just vague and no one knows what the problem is or no one has specified enough. But in mathematics, there are well-defined problems for which there's no general method for solving all instances of it. So in school you may learn how to solve a certain kind of equation, and then they teach you how to solve different instances of the equation. But if you think about complex enough sets of equations, then provably, there's no method… There's no method for solving every instance of it.

McNally:

I was going to say, and it seems to me this is the territory, the foundation, if you will, for where your concept or your book Probably Approximately Correct, sort of comes into play, right? And what is the advance that that offers?

Valiant:

Well, just like… I suppose… Turing specified what it means to compute. So after that, there was nothing ambiguous. You had a computer program in some rich enough notation, everyone understands what it is.

So this probably approximately correct model of learning, again, the specification of what you want to achieve, to be able to realize it. For example, if currently with large language models, they're trained in the first instance to the next syllable. So that's an instance of this is taken many examples, you an algorithm which produces another algorithm which can predict the next syllable.

But the important thing is that there's also some quantitative constraint that you want to be able to make these predictions without having taken more time than there has been available in the universe to do this. So you have to be able to do all those things, both the evidence you need, the number of examples you need, and also the amount of computation you do. They both have to be practical. They have to be…

McNally:

Right. I thought that was so interesting when I learned that, which is that we've started with solvable, unsolvable, but something could theoretically be solvable. But if it took infinity to run the program…Your PAC learning model says there's a line before that.

Valiant:

Right. And it doesn't even have to be infinity. There's lots of things which are very expensive, like exponential time. So there many ways things can be impractical, be large, doesn't quite have to be infinity

McNally;

And what's going on in AI right now. We hear a lot of talk about the amount of energy it takes, the amount of servers, all of that sort of thing. Two questions. One, I think what that tells me is that what we're doing would've been impossible prior to our technological innovations. Our technological development has allowed that. And secondly, just some thoughts from you about this notion of how much it takes to do ai and will we get more efficient at it?

Valiant:

It justifies the PAC model, which concentrates on the costs of doing learning. And if we want to learn better, then probably we need to do even more computation. But of course, anytime we do some computation, we can aim to do it more efficiently or have a better idea. So there's always hope that whatever takes a long time now we can reduce the costs.

So it's always some progress in doing things more efficiently, more economically, and maybe just doing enough for what we want to do and not doing too much. I think the present tension is that if we want to have systems which are more capable than now, then the obvious direction to go is to throw even more energy at it. But clearly what we ought to be doing is to find more efficient methods. But the evidence is that there will continue to be this struggle between the two, between functionality and the costs trying to blow up.

McNally:

One of the things that I also found interesting was that in terms of PAC learning the conclusion, tell me, and as I sort of said at the top, this particular conversation is intended for layman and being conducted by a layman, no question. But that PAC learning that says that averaged over many cases your predictions will be good, but over any single case, the promise is weak. And it seems to me that has a lot to tell us about our use and our dependence on ai.

Valiant:

Well, yes. Yeah, I think you've got it exactly right. That in the first sense of doing, if getting each prediction right isn't crucially important, we just want to do well on the average, then we've got a fantastic technology. But if you're looking, there's one decision you have to make and it's life or death for you, then you have to think, you have to think very carefully how you use AI or how you use anything. If you ask people's advice, what you should be doing, again, you have to be very careful.

So AI has not solved that problem longer, long as the framework is like this SPAC model, which I think at the moment it is that it's very good on doing things, doing many things on a large scale and being accurate on the average. But it guarantees any single instance are extremely weak then. So we're left with a dilemma exactly as you stated.

McNally:

And I have a feeling that most people don't realize that they've heard about hallucinations, which for anyone who hasn't read up on AI is that AI will go out and get you something that sounds really good, but it may have factual errors in it that you won't see because it sounds so damn good.

Valiant:

Yes. So in fact, I'm also a bit amused that the word hallucination as used is a bit generous to the machine, because when you hallucinate, the idea is that you really know what's going on, you're just making some sort of minor mistake or something like that.

McNally:

This is almost a feature, not a bug.

Valiant:

Right. And hallucination, as you're suggesting, is not the right term for it. It's being too generous to the machine.

McNally:

A lot of your thinking and your work focuses on learning…when we go way back to your early decisions of where to go once you discovered the field and so on, why is that? Why is learning so fascinating to you and compelling? And what should people know about learning that they might not know when they hear the word?

Valiant:
Well, I think part of my history in computer science is that I decided fairly early on that learning is kind of the fundamental part of AI. When I started, AI was various topics, and it wasn't clear which one would be the most fundamental part. So some people emphasize natural language, some people emphasize reasoning, some people emphasize planning. These are all valid things to study. But I thought there must be one which is the most basic one, and I decided that it was learning.

I think the most obvious alternative would be reasoning. Obviously reasoning is very important. But again, no one has found a kind of absolute definition of reasoning. I think what we do is we learn different ways of reasoning. I suppose in school we learn to do arithmetic. It's like a special kind of reasoning. And we get good at it because we see lots of examples of it. And then we learn some other methods of reasoning.

Logicians, of course, for thousands of years have been trying to figure out - what is reasoning? what is reasoning? But I think they haven't come with a totally satisfactory answer of some single system. The early AI systems, the idea was that we'll put in lots of knowledge, we'll put in some method of reasoning, and see what happens. And it kind of didn't work out for various reasons.

My theory is that our most basic relationship with the world as humans, and same with other species, is that we can absorb information from the world, make observations in the world, and from those observations, we can generalize and make predictions what will happen in the future. And this is roughly how we cope with the world.

So this is the basic problem that we see. We see examples of the world, but then we have to operate in situations which aren't exactly what we've seen. And so learning is basically in this formulation, is the ability to be able to perform well even in situations which we haven't seen, as long as they have some similarity to what you've seen. So if the world changes totally suddenly, then you're helpless, but…

McNally:

Right. In other words, too radical a change, too enormous, it won't work, but if you can make leap to leap to leap…

Valiant:

Exactly, exactly. And I think this basic ability to learn from examples has been there from the earliest days of evolution. Very primitive creatures figured out how to feed, where to go to, how to avoid predators. And they could adapt to the world, they could make observations and do sensible things. So without that, without some method of adapting to the world, adapting to your sense of the world, to your observations in the world, one wouldn't be very fit.

McNally:

Yeah. Going back to that word.

Valiant:

One problem which evolution solved, I think, is that living things can adapt to the world.

McNally:

Now you make the point that when we try to say what makes humans unique, when we try to say, why did we succeed, and so on - that the more and more it seems to me from Survival of the Friendliest and the work of David Reich and folks like that, we're learning more and more that animals can do an awful lot of what we thought…

It seems to me you've got to throw out every few years lately what we thought was the thing that we could do that they couldn't. And then it turns out crows can do it, and then it turns out octopuses can do it even better, and all of that sort of thing that we've been learning. What makes educability pass muster, at least today?

Valiant:

Okay, well, it's a great question. I think my answer to this question of what makes humans different… The answer isn't entirely trivial, okay. So no single word in the English dictionary, no single general idea is enough, because people have tried them all and they don't work - often because animals can do it as well, but also because by themselves they don't explain that much.

My work is a bit more kind of technical than most of the other approaches in this business, but I think the point of being a bit more technical is that you can say things a bit more precisely. And because you can say things a bit precisely, you can say things which are slightly more complicated.

My definition of educability probably takes a paragraph to say, and I think the solution to the problem is something which takes about a paragraph, and the single word solutions or single face solutions don't work. But the other thing I should emphasize, and what I'm trying to do, which may be different from what other people are doing in this business, is that I'm trying to explain kind of the cognitive capability.

So what's our cognitive ability which gives rise to all these things which humans can do and other species cannot? You say there are many theories of what's special about humans, but I'm suggesting that there's very many approaches… Language is one many people emphasize… social learning…

McNally:

There was tool use for a while, but we found out that didn't work and so on. And the social communication, the ability to work collectively, those sorts of things…

Valiant:

That's right. So I think there's an infinite variety of things which humans did, which were important in our history, but aren’t the kind of explanation I'm looking for. So I think that if we figure out what the basic cognitive capability we have which is different from animals, then all these different other things are probably …a consequence of that…

Animals have their own language, but this incredible flexibility we have with language probably is a consequence of this more fundamental capability, which I call educability.

I think I'm facing the question a bit differently from many of these other workers. I want something a bit more technical, a bit more precise, but something which may almost be at a slightly lower level than what other people talk about. I think at this fundamental level of what's our cognitive capability, we're trying to explain how these other capabilities we have, arose. People say that animal domestication, that's very important in our history, There are many other things which were important in human history… to the history of civilization, which apes can't do, and I'm trying to find what this common capability is, which enabled us to do all these many things we do.

McNally:

Right… Am I correct? I think I heard you sort of saying that if you can get something that is both precise enough and useful enough that you might then have created sort of a foundation or a fundamental distinction, which then those other distinctions which people have been noticing relate to or is built on, that sort of thing.

Valiant:

Yeah, yeah, exactly what I'm hoping for. So the level I'm trying to work on… I think there's a fundamental problem. I'm suggesting a solution. Other people may want to change it, but I think that's a fundamental thing to do there at that level. I hope that in terms of however that resolve itself, one can go and explain more complex phenomena,

McNally:

Right… that even if what they end up doing is using your proposition as a tool for further exploration, you're good with that?

Valiant:

Sure, sure, sure.

McNally:

It's not about having the final answer, it's about building the knowledge, right?

Valiant:

Yes, yes. But also that with complex phenomena, there are explanations at many different levels, and I'm trying to supply one level.

McNally:

Okay, let me tell people that this is Free Forum: A World That Just Might Work. I'm Terence McNally, I'm speaking with Turing Award-winning Harvard Professor Leslie Valiant about his new book on what enables humans to thrive, and the dangers we face, *The Importance of Being Educable: A New Theory of Human Uniqueness*.

Let me just go into this educability just a little bit more. Contrast it with what people sort of generally think of as intelligence or smartness or that sort of thing.

Valiant:

Yes. So I think the main problem I have with the phrases intelligence and smartness, is that there's no definition of what it means. That is not actionable. So you can say, "How do I recognize whether someone is intelligent or smart?" You may say that you see it, that you can tell when you see it, but I'm hoping for a more actionable definition.

So, for example, if you want to make computers to do useful things, then I think the word artificial intelligence has not been useful at all because people didn't know what intelligence meant, but when we start talking about learning, then, say, learning from examples, it's pretty clear what you're talking about here. You get examples and trying to learn to do something else.

It's like learning from examples is clearly… one can say more formally, it's clearly a specification of what you want to get done, whereas intelligence is not. And I think also historically, IQ tests came from the idea of implicit definitions or just correlations. So people observing that children in schools… Different children did differently well in different subjects, but some did more better in many and some did worse in many. So people hypothesized that the ones who did well in many, they must be intelligent. They hypothesized there's some commonality, but these are pure implicit hypotheses. They never told us what it is that these children have in common.

McNally:

I must say that I think growing up, one is interested in IQ. "How am I doing? Where do I fit in?" All that sort of thing. And yet I think it never made sense to me that a test that's based in a certain culture and a certain era, and all of that, could really be tapping or calculating something as ineffable as intelligence.

Valiant:

Well, exactly. I think there you have it. Yeah. But I think exactly as you say, it's ineffable intelligence. So I think a more useful direction is to define capabilities more precisely, and then maybe you can test for them.

McNally:

If you define capabilities more precisely, then you're getting away from this ineffability.

Valiant:

Yeah, yeah. So, for example, if someone gives you a test in Spanish language, that's not objectionable because it's clear what they're testing you for. But testing you for something where it's not clear what they're testing you for is a bit objectionable.

McNally:

Yeah. Very good. So let me just jump a little bit and ask you - How did this book happen? I'm often interviewing authors, having conversations with authors, and I'm almost always curious… why this book at this time? what was the first itch you scratched? When did you say… because you've been prominent in your field for a long time, and this is the third book…?

Valiant:

Right, right…Okay, so in my field, we usually write papers, not books, but this needed a book. In some sense, it was a culmination of a long time thinking. Learning is part of educability. Then for long time, I thought about how to add reasoning, given that a machine has the capability of learning from examples. What does it mean for it to be able to reason in addition?

Okay, so I worked on this for a long time, and came up with a system called Robust Logic, mentioned in the book. I reconciled learning and reasoning in a reasonable way in my mind, but I still knew this wasn't quite the answer. I thought one should be able to do better, that if one just made a machine, which did just this, it still had something missing.

So this notion of educability basically adds a third thing to it, which is in some sense, the simplest, which is the Turing idea is that besides being able to learn and reason in these senses, you can also execute programs given from the outside by instruction.

I think the most primitive thing is to have experiences and using things from that. But we humans, I can also learn from your experiences. You have your experiences, you make a deduction, you make some conclusion about what the best thing is to do in a certain circumstance, and you just tell me it. And I benefit from your experience.

And of course, this is how science progresses, that people do scientific experiments with great labor, but then if they get good conclusions, they can tell a thousand people in a lecture room in half an hour, and they don't have to repeat the experiment. So the idea that you can transfer from one person to another, the conclusion. The conclusion could be a computer program, a recipe, a formula…

McNally:

I would imagine even a new way like the Fosbury Flop.

Valiant:

Yeah, sure. Exactly. Exactly.

McNally:

One way of high jumping came… suddenly someone…and within a couple of years, no one did the old way that they'd done forever.

Valiant:

And clearly this is critical for human civilization that various discoveries made with great difficulty, once they were made, a thousand people could copy them the next day. I want to combine these three things - of being able to take instruction, being able to learn from experience, and the third thing is being able to reason in a certain way.

So reasoning here just means being able to chain together the various things you've learned. If you respond to something I'm saying, you may be chaining together things you learned 30 years ago from somewhere, from your experience, something you learned yesterday from a book. And you can freely apply the knowledge you've learned at different times to draw a conclusion now.

McNally:

I listened to a podcast you did with Sean Carroll. And you make the point that it's the integration of those three things, right? In other words, some of what you were saying sounded like that ability to socially communicate and bring people together and work together to solve problems, but there's a nuance that's different in what you're saying, isn't there in terms of how they integrate?

Valiant:

Yes, because, so I'm talking about our individual capability. What capability do I have? What do you have? And I think one application is that we can collaborate and do something together, but that's probably something pretty complicated. And to be able to do that, we use all these abilities we have in our heads of being able to learn from experience, reason, and being able to take instruction from each other. I'm saying that this more complicated phenomena. So the social phenomena in humans get obviously complicated…

McNally:

I think what I just realized is that the ability to socially communicate and bring people together to collectively solve problems may be a description of what we do in groups. But you're actually looking for what in each individual allows us to do that in groups.

Valiant:

Exactly, exactly. That these complicated things we do in groups require a lot from us as an individual in our cognitive actions. I think that's the fundamental part.

McNally:

You took it one level down in a way or one level up, but to this - What capabilities do individual humans need so that social integration is possible?

Valiant:

Yes, yes, yes. At the level in which humans do it. So clearly other species can do social things to a certain extent, but they're not as successful as we are in doing and building complicated things together.

McNally:

Let's talk a little bit about how this relates to AI, large language models, and so on. And I think one of the things I appreciate that you said about large language models, was that they've been designed to, and what they do well, is predict the next syllable word, token, whatever you call it. And you said that any other abilities we ascribe to them, like "Hmm, sure sounds like their reasoning… Sure sounds like they're integrating things," is us. It's projection. It's intuition.

Valiant:

Yes. So of course, I mean now large language models have changed a bit since I wrote the book. So now people do try to add explicit things which look like reasoning, but sure, I think so. My main reaction to the large language models was how much people intuit into it. So just having smooth natural language production sounds so human-like to us that we really intuit that there's a human in there. So our reaction to large laguage models is worthy of study by itself.

McNally:

But it seems to me one of the fun things you said, besides the fact that I just love that when you said, "…and anything else is probably intuition…" And I'm going to ask you in a second about how rapid the development is that you're saying that that statement might not be so true today. But the other thing you said was that, instead of feeling like, "Oh, I thought they were reasoning," you say it's probably worth valuing the fact that they can fool us. In other words, the fact that we imagine their reasoning is an achievement of large language models in itself.

Valiant:

Well, yeah, I'm not sure whether I said exactly that, but certainly this is what we find out from AI is that we find out both what they're good at and also what we are bad at.

So if AI systems can beat us in Go, it may just be that humans aren't very good at the game of Go. And similarly, we may not be able to evaluate too well. If we hear some plausible text, we are probably not very good at evaluating whether what it's saying is plausible or true. I think that's a certainly the case. Yes, yes.

McNally:

Well, what you're saying is that the weakness, the pitfall that I referred to at the start, which my guess is, and you tell me, but I was thinking that might've been part of what prompted the book at this time as well.

Valiant:

Well, actually it wasn't. No, it was purely the science. These two things came together and I did decide to write a general interest book. I thought the whole concept was of general interest, and then I did try to see what are the consequences of this?

McNally:

Right, right. Okay, I get it. So in other words, the impetus to, "Okay, this is a book," is the realization of the unique capability and the integration, all that. But once you're sitting down and writing a book, then the consequences and the lessons become important.

Valiant:

The obvious consequence, which I was surprised by, is that with all these capabilities I described, learning from example, chaining together, knowledge, taking instruction - nowhere is there a place for us to check whether something is true.

So we are very good at manipulating knowledge in all kinds of ways. We can absorb knowledge. We absorb knowledge all day, we read stuff. If you tell me something, it reminds me of things I learned years ago. We are very good at running with knowledge and relating different pieces of knowledge, but nowhere is there a capability which is there to check whether something is true. If you tell me something, my first response isn't some capability to tell whether it's true or not.

When you learn from experience… you learn where the best fruit is or where to avoid predators, and the world isn't changing, then what you're learning is kind of reliable, that you can have faith in it. No one is fooling you. You're in touch with reality. And even if you reason these things, you're still in touch with reality.

But once you start absorbing knowledge… you read a novel or you hear a podcast, who knows whether what you're hearing is true, worth knowing, or relevant. So in this sense, I think our species is in a new position compared to other species.

McNally:

There's been a lot of engagement over the last few years with cognitive biases, the work of Kahneman and Twersky, and all of this and so on. And I think it's been very fruitful. Now, if you actually look in the real world, you'd say, is it being actually put into use, if you get what I'm saying, there's an awful lot of people following mis- and dis-information. Can you talk about those and how our susceptibility to propaganda and manipulation is related to what you're doing and related to these biases, how that all fits together?

Valiant:

Yeah. I think, where there's a slightly new angle on this from this educability notion is that it does focus more on our weaknesses. That much of the discussion is that there are some bad guys who tell us false things, or there are new methods we could be fooled by, new technologies. Everything is true and everything raises problems. But I think there's also scope with educating ourselves into understanding how we're… understanding our basic weakness.

This field of propaganda, the ways by which humans can be influenced… advertisers, political propaganda. So it seems to me that it's in everyone's interest, at least in a democracy, to have the population really well-educated, to be alert to these, I would say, threats. So if there's more and more information being circulating around, surely it’s in everyone's interest that humans should have some protection, some self-awareness of what they're subject to.

I think the slightly new angle, which I thought was different from what other people were saying, was that we should educate ourselves into understanding this very basic weakness we have - that our strength of being able to absorb information all day and run with it and integrate it has also this weakness…

McNally:

And you talk about we learn in environments - and the environment can be helpful, neutral, or adversarial.

Valiant:

Yeah. Yes, exactly. This basic learning process… You can investigate how it does in each of these environments. If you're teaching a small child, then presumably the environment is trying to help - so nice examples, doesn't lie, et cetera, et cetera. Much of our experience is kind of neutral. If we walk around, we just observe the world and no one is trying to trick us into. But then obviously there are also adversarial situations, where someone is trying to persuade us to do something, and to persuade us to do something when we are not even aware that they're trying to persuade us.

McNally:
And you say that a particular phenomenon is gaslighting?

Valiant:

Yeah. Okay. So that's an extreme phenomenon. Yeah.

McNally:

Takes it to almost another level where it's not just this fact might not be true, but this whole package of stuff might be true to where now you don't know what's true.

Valiant:

Yes, yes. So the threat isn't just that someone tells you a single lie, but that they give you information which totally changes your perception of something beyond just a single fact.

McNally:

You make the point that in a free society - which basically we are - the prospects of controlling propaganda have limits because of free speech and things like that. What's your hope, your vision of education that could actually deal with the weakness of our educability?

Valiant:

Well, I think the main thing I would just repeat what I was saying before is that, so some people do study propaganda, ancient Greeks studied rhetoric, which is how you influence people by speeches. So presumably if you studied that, you both learn how to influence other people, but you also learned how not to be influenced by other people who are trying to influence you, but you didn't want to be influenced.

As far as education and civics, besides just learning the mechanics of how a constitution works, it seems equally important to learn the weaknesses an individual has when subjected to all these many forces of persuasion in such a democratic society.

I think giving greater emphasis to education which highlights our psychological weaknesses and being easily misled. So I think that's the main thing. I mean, the other kind of area which I think one can think about as far as action items is that they do exist institutions, which even in a society which is politically very conflicted, is so constructed that maybe if there are two sides of the conflict, that both sides contribute to the construction of this institution and therefore have a stake in accepting its conclusions.

So for example, in the Congressional budget office. Congress is a political organization, but they figured out some way in which they can agree on certain financial things. Having institutions which are constructed so as to have more respect through society, I think is something one could also encourage.

McNally:
Now, and I'll sort of make this my last question - It seems to me that in the current moment in the United States, and it may be happening in other countries as well, there is discouragement, disinvestment in education at the lower levels, and certainly higher education is under attack. So it seems to me where we're going to be able to do more sophisticated education, which is what you're talking about, is challenged by the fact that it's getting harder just to do what we've been doing.

Valiant:
Well, yes. I mean, if that's true, that's certainly very sad. But yeah, so I think if you looked at the book, I do believe that education is even more fundamental than I thought before I wrote the book. So I think it's very fundamental to who we are. And you…

McNally:
And you're someone who's been thinking about learning for decades.

Valiant:
So I think if education is the most fundamental thing about our mental life, I think, and certainly the challenges today of AI, everything else, we will have to rethink what we do about education. I think devoting a lot of energy and resources to thinking about what humanity should do about education in the future, I think it's something which should be a number one priority.

McNally:
In other words, this book sets out a new way to look at things, and sets what you just did as a primary goal. If we're going to succeed in the coming decades… and let's just put climate change, for instance, over here. We've got problems we need to solve, but this is the early work. This is not, "Here's five examples of people doing it well," and "This is the theory of how to do it." You're just saying, "Here's the charge."

Valiant:
Yes, you are right. That's the main summary - that we should put a lot of resources into having new thoughts about what to do about education.

McNally:
Okay. Okay, we're going to leave it at that.

So again, the book is *The Importance of Being Educable: A New Theory of Human Uniqueness.*

For this conversation and many other interviews and articles, to join me in pursuit of a world that just might work, go to terrencemcnally.net or aworldthatjustmightwork.com. They're the same website. If you want to get my weekly announcement telling you who's going to be on what we're going to talk about, and usually links to 10 or 15 articles that flesh out the conversation, email me at temcnally@mac.com

You can find years of podcasts at my site or at Apple Podcasts, Spotify, all the major podcast sites. Michael Lewis, Jeremy Scahill, Naomi Klein, Robert Reich, Van Jones, Connie Rice, Greg Boyle. You can follow me on Twitter @mcnallyterrence. And thanks to Kiyana Williams in production, George Vasilopoulis at Progressive Voices, and most of all, to you my listeners. Please share this podcast widely.

And finally, thank you, Leslie Valiant. Keep up your good work.

Valiant:
Well, thanks for inviting me.