Dr. Winston Ewert: The Limits of Artificial Intelligence

https://mindmatters.ai/podcast/ep331

Pat Flynn:

Hello everybody and welcome back to the podcast. This is your host, Pat Flynn. I am once again teaming up with the amazing team at Mind Matters, and we are discussing another contribution from the volume Minding the Brain. This one from Dr. Winston Ewert. In Part One, we have discussed his position on human cognitional ability and it essentially being just a very, very sophisticated algorithm. Now, at first you might think, "Okay, well isn't this what a lot of atheists, materialists and people who hold to computational theory of mind say? What does that mean? Does that mean that that world view is true and that there's nothing of any huge grand significance to us?" And we're about to say not so fast, not so fast, because in part one, we differentiated between different aspects of the human mind. The phenomenal consciousness aspect, the what it is like aspect, the what it is like to taste chocolate or to hear a symphony. And Winston, in agreement with many of the other contributors to Minding the Brain said, "Yeah, that's definitely not just a formula or an algorithm."

But when it comes to our cognitional ability, our problem solving ability, well, there's more of a case to be made there. And indeed he made that case. So please listen to Part One if you haven't already. But the implications are not so great for many people who hold to a reductive materialist world view. And we're going to explore that right here and right now. So Winston, thank you so much for taking the time to be here.

Winston Ewert:

Yeah, it's good to be here.

Pat Flynn:

Okay, so, all right, where should we begin here? Because a few things I want to explore, and obviously they all concern the implications of a lot of what we discussed in Part One. I want to explore AI, but I also want to explore why you think this understanding of the human mind is actually pretty bad news for people who think that you can reduce the human person to atoms and the laws that used to combine them and Darwinian forces and all that sort of stuff. So yeah, help us find a place to start, Winston.

Winston Ewert:

So I think the key thing is the linchpin of my argument is the idea that an algorithm could only construct an algorithm less sophisticated than itself.

Pat Flynn:

Ah, yes. Okay, yeah, help us understand what is meant by that.

Winston Ewert:

So we've got the idea of the halting problem. So my argument, all kinds of cognition can be reduced to examples of the halting problem, including the problem of constructing a halting problem, a halting detector. And so basically it means if you wanted to build an artificial intelligence that's equivalent to building a program that's able to determine whether certain algorithms halt. And that itself, you could

build a program that verifies the correctness of a halting detector by simulating, checking through all of its programs trying to find one that it gets wrong. And if you can't find a program it gets wrong, then just like when we were searching for a counterexample to Goldbach conjecture, we check to see whether there's a counterexample to Goldbach conjecture. We could also check for a counterexample to a halting detectors claims.

Now the key thing is that the fundamental proof we have in computer science that you cannot build an algorithm which solves all halting problems, basically shows actually that no halting detection algorithm can correctly detect whether itself will halt. So basically what I argue from that is, well, basically a halting detector can only detect whether a less sophisticated halting detector is correct because if it could detect itself or a more powerful halting detector, that runs into a contradiction in the standard proof for the impossibility of halting detection.

Pat Flynn:

Okay. Really fascinating there. So my first immediate question is, assuming this is well known among the relevant experts, why do so many people seem so optimistic about singularity? If you want to explain what that is for us and all the, essentially Terminator type scenarios and all that fun stuff.

Winston Ewert:

So the singularity is the idea that eventually you are going to have an artificial intelligence smart enough to create an even smarter artificial intelligence. And then that will create an even smarter artificial intelligence. And there'll be this breakaway explosion of artificial intelligences that far surpass human ability and leave us far in the dust and change the fabric of the universe forever.

Pat Flynn:

Yeah.

Winston Ewert:

That's the basic idea of the singularity and an implication, my argument is that doesn't happen because if it-

Pat Flynn:

It can't happen, says your argument.

Winston Ewert:

It can't happen because an algorithm can only drive. Now you ask if this is well known. So what's well known is the proof that you cannot develop an algorithm which correctly classifies the halting status of all algorithms. That's well known. The idea here that, the implication is that an algorithm can only construct a less sophisticated algorithm, that is basically original to me. So that's not a well-known conclusion. You can get, in the chapter I talk about, I'm drawing a blank on the guy's name.

Pat Flynn:

Was it Penrose?

Winston Ewert:

Yes, Penrose. There it is.

Pat Flynn:

There we go. Always happy to lend a hand.

Winston Ewert:

Penrose has actually a somewhat similar-ish argument that he makes, but he draws a different conclusion than me from it. So he realizes the problem and says, "Well... His solution is that the human mind must not be computational.

Pat Flynn:

Sure, yeah.

Winston Ewert:

And so I take a different fork there on his argument and say, "Well, it is computational." The problem actually is another part of Penrose's argument was specifically that, well, the only way this could work is if humans had a really sophisticated algorithm and where would that come from?

Pat Flynn:

Yes. Right, and he's not willing to hold, of course, to a divine origin or a divine intervention.

Winston Ewert:

So a divine origin of course, solves a different problem, solves the problem a different way. And so it's similar there. But in general, there's not a widespread awareness, I think, of this particular issue.

Pat Flynn:

Okay. So how would you draw it out for people who are familiar with the relevant proof and you want to try to convince them of taking, I guess either your or Penrose's option? What's the simplest way to make that connection?

Winston Ewert:

So the basic proof for the impossibility is they say, well, imagine you take your halting detector and you add an addendum to it that says, "Okay, if I would halt, if the program itself would halt, then loop forever. And if I would loop forever, then halt." So it's a contrarian program that says, "Whatever I think I should do, then I'll do the opposite." And so that of course leads to a contradiction because either it got it wrong about what it would do, in which case it's not correctly detecting the halting status or it can't do the opposite, which it's a very trivial algorithmic step to do the opposite of what your prediction claimed it would. And that conclusion is easily avoided if you say, "Well, I can't run on myself." And that's really the only way to avoid it. And so you can construct an algorithm which, in principle, could detect everything else except for itself, and that would be consistent with the proof.

Pat Flynn:

Ah, yes. Okay, I see. All right. So is this strong enough to say that you really can't, at the end of the day, have, and I know you talk about this in your paper, gradually increasing intelligence?

Winston Ewert:

Right. So I think it does rule out these gradually increasing intelligence because if in order to build a more sophisticated intelligence, you somehow have to know that what you're adding onto it is correct. And how do you know that if you can't evaluate the correctness of that already? And if you could evaluate the correctness of that already, well then you're already that intelligent.

Pat Flynn:

I see. I mean, ultimately your argument, a master argument against the possibility of the singularity, but also a very reductionist evolution, neo Darwinian evolutionary explanation of human cognitional ability. Is that right?

Winston Ewert:

Yes.

Pat Flynn:

Yeah. So the idea there is while God could of course create beings of increasingly gradual intelligence, you don't have ultimately gradually increasing intelligence because you have the ultimate intelligence at the bottom of things. Is that right?

Winston Ewert:

Yeah, that's the basic idea.

Pat Flynn:

Yeah. Okay, that's a really fascinating argument. What I would like to maybe just explore towards the end here are just as you usually do with stuff like that, it's provocative. This is interesting. Certain objections. What objections have you found so far to your work along these lines? And apparently you still hold the same position, so why do you think they fail?

Winston Ewert:

So I haven't had a lot of people take notice of the argument I laid out here yet and try to-

Pat Flynn:

Oh, they will now, so get ready.

Winston Ewert:

Well, I got to be ready for that.

Pat Flynn:

Well, one thing I just want to point out that I like, well, maybe we could even phrase it this way, and certainly happy to hear any informal objections you may have heard too, but what I like is your argument makes a prediction. If tomorrow the Terminator systems come online and we find ourselves in a singularity, I guess you'd be changing your mind, right?

Winston Ewert:

Yes. I mean, that would be the least of my concerns at that situation, but...

Pat Flynn:

I've been refuted now. There's bigger concerns. Are there any other predictions you think that your particular understanding might, whether with respect to AI or anything in general?

Winston Ewert:

I mean, I think it predicts that we're going to see increasing sophistication of AI because basically it says the prediction is that humans can't create something more intelligent than themselves, but they certainly should be able to create something increasingly close to themselves as they work on it over time.

Pat Flynn:

Sure. Yeah, that makes sense.

Winston Ewert:

And so I think even since I wrote this chapter and ChatGPT came out and we've had a bit of an AI revolution, I think that that prediction has actually turned out fairly well.

Pat Flynn:

Oh, wow. So you wrote this even before Chat really had a-

Winston Ewert:

Yes, ChatGPT. It was a bit of a thing that just how long it takes things to progress through the publishing process, is like, "Wow."

Pat Flynn:

Yeah, I understand, yep.

Winston Ewert:

I mean, everything in the books obviously now that ChatGPT has come out, but I think that the stuff we have in the book stands up well to-

Pat Flynn:

Yeah, it's neat to be able to see that because I know your contribution isn't the only one that was on the other side before ChatGPT really exploded onto the scene. So it is cool to see that now that the book is out and ChatGPT is pretty much being used by everyone to see how well a lot of these predictions and anticipations have stood up, to be sure. Is there anything about AI now that has really surprised you at all, or that maybe you think is a tension for your model? Or is it all pretty much accommodated by how you're thinking about human cognitional ability?

Winston Ewert:

Yeah, I mean, I think that the abilities have, I mean, I'm impressed by what ChatGPT can do, and I was somewhat surprised, although I had seen, I think a GPT 2 model and already come to the realization of, "Okay, this technology is better than I thought it would be at this point in time." And that was before I wrote the chapter that I would've seen some of the precursors to ChatGPT that maybe didn't get as

much press. But I think that in general, I feel it does fit very nicely into the framework that I was thinking of in terms of we expect the progression, but we think there are limits to it.

Pat Flynn:

That there are limits and limits where we would presumably be able to know if they were surpassed, right. And I think that's a neat feature that yours has those specific sorts of predictions as well. All right, very cool. Is there anything else about your article or your argument that's important that you think we should mention? Obviously, we're going to encourage people to pick up Minding the Brain and read it if they want to dive into the more technical details, of which there are many, we're just mostly trying to get people interested. But yeah, anything important you think we missed or glossed over?

Winston Ewert:

I think maybe one thing to think about, it's worth emphasizing is the argument is, of course, that I make, is that an intelligence can only create a lesser intelligence than itself. And so the implication then would be something more intelligent than us must have created us. But there's an interesting hitch there because, well, we can't just postulate an infinite regress of more and more intelligent algorithms responsible for us. So at some level you have to break free of that and say, "Well, whatever designed us ultimately must not have been an algorithm. It has to be something that is somehow transcendent and beyond algorithmic status," which I think is very, very suggestive.

Pat Flynn:

I think that is very suggestive indeed, especially of the traditional understanding of God in line with classical theism and this or that. So that's always great. But you're not engaged in a full project of natural theology here, so...

Winston Ewert:

No, but I think it's definitely very interesting that it points very firmly in that direction as the explanation.

Pat Flynn:

Yeah, because what you wouldn't want to get to, presumably is just some brute immense algorithm, right?

Winston Ewert:

Right. I mean, because then you'd have somewhat the same problem. Where did that come from?

Pat Flynn:

Exactly. So what you want is something that can explain the algorithms without itself being an algorithm, right?

Winston Ewert:

Yes.

Pat Flynn:

Yeah, I'm totally on board with that. That sounds great. Dr. Ewert, thank you so much for your time here today. Before we say goodbye, two questions. What are you working on next and where can people keep up with you and your work?

Winston Ewert:

Well, people are familiar with me will know that I recently published the second edition of the Design Inference with William Dembski. I'm currently working on a follow up to that called Get Your Own Dirt.

Pat Flynn:

I like that title. That's fun.

Winston Ewert:

It's a reference to a joke that you've probably heard about a bunch of scientists go to God and says, "We don't need you anymore. We can create life." And God says, "Show me." And so the scientists go and grab some dirt and God says, "Wait a minute, get your own dirt." And so it's playing on that, making a design argument that says even if you're going to claim the world is such that evolutionary processes or things like this actually work, that just pushes the problem back and you're getting your dirt from somewhere. So that's one of the major projects I'm working on next. I have a website, winstonewert.com. You can go there, see links to my various projects and papers.

Pat Flynn:

Excellent. Dr. Ewert, it has been a pleasure. Thank you so much for your time today.

Announcer:

This has been Mind Matters News. Explore more at mindmatters.ai. That's mindmatters.ai. Mind Matters News is directed and edited by Austin Egbert. The opinions expressed on this program are solely those of the speakers. Mind Matters News is produced and copyrighted by the Walter Bradley Center for Natural and Artificial Intelligence at Discovery Institute.