AI and You

Transcript

Guest: Alison Gopnik

Episode 97

First Aired: Monday, April 25, 2022

Welcome to episode 97! Today I will finish the interview with Alison Gopnik, who is the American professor of psychology and affiliate professor of philosophy at the University of California at Berkeley. She works in cognitive and language development, specializing in the effect of language on thought, the development of a theory of mind, and causal learning. Her has appeared everywhere from *Scientific American* to the *Wall Street Journal,* she has given a TED talk, and been on TV shows like The Colbert Show, and she is an expert in how babies think.

Last week we learned why that is so relevant to AI, and how her research, which shows that babies are a lot smarter than I used to think, that they're really little scientists, informs our progress towards artificial general intelligence. We talked about some of that research, how babies are even surprisingly good at Bayesian – or probability-based - calculations, and her involvement with DARPA's Machine Common Sense project. Now let's wrap it up as we get into topics like epigenetics, and the AI alignment problem. Let's get back to the interview with Alison Gopnik.

Are you able to differentiate between what a baby has learned from its environment and what came with its DNA, what it had biologically?

That's of course also one of the great big questions, because certainly the idea that a lot of this is built in is one of the approaches to solve this. In general, I think in developmental psychology we've increasingly recognized that that distinction is also not a distinction that's all that helpful. So the right way to think is that we have a creature that's obviously has a genes, has DNA But even in the womb even when the brain is getting wired up you already seem complicated relationships between the data that's coming in the environment, the external world and the way that say that DNA Is building connections in the brain. So it's the nature of the way that biological systems work is that this interaction between what's there to begin with or better way of putting it is what's there in the genes and what's there in the environment, that's just a constant cascading two-way interaction from the moment you have a fertilized egg maybe even before the moment you have a fertilized egg. And the learning pieces just one piece in that very complicated back and forth interaction. And what we're trying to do is to figure out ways of describing and characterizing exactly how that happened. So for us in this kind of problem specific pieces of the environment that we're concerned about, what's the pixels that are out in our visual world for example. And how does that lead to things like our representations of objects. But that's not a difference in kind from how is it that a piece of genetics is expressed in a particular context in order to make a particular protein for example.

Do you get into how it might go the other way? I think, revealing some ignorance, the word for this is epigenetics. That when we learn things from the environment that it ends up in our DNA, assuming that's a thing?

So what does seem to be true again, this is this point about epigenetics is a is a striking example of this. But there are others as well that this cascade between the genes and the environment is always going in both directions. So things that are happening in the environment, whether it's sort of the physical environment of, for instance, the other neurons that are next to the neuron that is connecting or whether it's the environment of someone going out into the world and trying to make sense out of the objects around them, those things are always going back and forth. Now epigenetics is a particularly striking example where you can show that early environmental effects for instance can have cascading effects further down and then some of those can even be inherited. It's a complicated story about epigenetics. Some of them seem to be able to be transferred to the to the next generation as well. So things like how genes are expressed is something that can depend and change a lot depending on the environment and of course it's how genes that are expressed that's the thing that ends up making a creature that has a particular kind of characteristic.

I wonder if there is some reinforcement there of things that many people learn that reinforce in the gene pool at large and form some sort of racial memory that maybe we are now also the product of what many generations ago in hundreds of thousands or further years ago learned over and over. Like, that's a sabertooth tiger--run.

I mean I think that's not very likely, but what we do know is that we have a kind of parallel mode of evolution, which is cultural evolution. And there's nothing mysterious about that, we know that we pass on information from one generation to the next and the mechanisms for doing that seem to be very foundational with something that we see even again, very young children are already imitating the people around them, paying attention to what the people around them do. That's not genetic evolution, but it's a very powerful kind of mechanism that lets us change our behavior in the light of new kinds of environment and that lets us pass on information from one generation to another. So you don't need any special spooky mechanisms, just good old fashioned culture and cultural evolution can do that work of passing on information from one generation to another. And one of the - when I was mentioning our MESS set up the third leg after model building and exploration is this social learning, which is how we pass on cultural information from one generation to another. And that's another thing that we're very interested in is what is how is it the children are doing that as effectively and quickly as they are. And lots of people have argued, I think with some justification that that's really the secret of whether it's success or not, I guess remains to be seen. But that's certainly the secret of human change. The secret of how humans have changed their environments as radically as they have for example.

Talking about that human difference, when my children were born--and I'm interested in AI right, so I'm looking at them growing up, and I'm thinking, "I want to see that point where they learn language." I want to see how does that happen? That this is to me is this huge mystery of the bootstrapping of language. And so I was watching really closely to see when that spark happened and failed miserably. Has anyone done any better at that? Because that's like to me the big difference. And absent things like, yes, crows and chimpanzees and elephants have a form of language, but there's still a qualitative difference that that happens and it just showed up, and of course they couldn't tell me how they did it. And so that was frustrating, but how - I know it's a huge topic.

In fact I started out my career and I did my PhD looking, it's funny that you mentioned this Peter, doing exactly what you were doing, looking at the very first words that children were using and using those as a guide for trying to figure out things about how they were understanding about the world and how their early language worked. So it was a great project. In my dissertation, I spent hours sitting on the floor with 15-month-olds just recording every word that they said in the context, in which they said it. And I couldn't solve the problem either. I think I got some insights into what was going on, but there's a whole branches of developmental psychology. And interestingly now some of the big data techniques for example have made it much less unwieldy, much I guess you could say more wieldly, it's they've made it much simpler in some ways to help address issues like that, whereas I had to sit there and you know, literally right down from the videos everywhere that the kids said now we have ways of getting really big data sets and doing lots of analysis. So someone like Michael Frank at Stanford is an example of a developmental psychologist who's done really beautiful work, looking at all the language that the kids are learning and then looking at what they say, and trying to put those two things together. And I think in some ways the answer is not that dissimilar from the kinds of work in understanding objects or understanding people. The children seem to start out with some ideas about how language works, but they're also constructing new ideas about their specific language and constructing grammars and trying to make sense out of the languages they hear around them.

So even if your work did not end up making things easier for AI, AI will end up making your work easier in in ways of track data gathering and analysis. I noticed you stopped your TED talk right after saying maybe we should be getting adults to think more like children and I was like, I wish there was another five minutes there. And so if we could improve something about our learning as adults, if we were to learn from the babies, what do you think that would be? And would there be a downside?

So again, I want to emphasize the fact that there are trade-offs and explore-exploit is one of the great classic trade-offs that goes back to, you know, the beginnings of computer science. So even though it sounds great to say, well the kids are designed to explore, like you wouldn't want the kids to be, you know, your CEO or your department head, right? There's lots to be said for those kinds of adult capacities to ignore distractions, satisfice, make reasonable decisions, do it quickly, be effective. And those things are really intention with the childlike capacities to just explore a wide space of possibilities, try out things that aren't going to work, bounce around the space of potential answers, try things at random, be noisy, be impulsive, do risky things. And you need both of those abilities to, as an adult actually, get along in in the world and you probably need more of the exploit planning abilities and you do the exploration partly because you get to do the exploration when you're young and partly because we have a kind of division of labor where we have people like scientists or artists who kind of get to do that kind of broad-ranging exploration. But I think one thing that I'm increasingly convinced of is, in order for that to happen, especially with children, a very significant piece that we don't pay enough attention to is we need to have caregivers, we need to have other people who are taking care of you while you're doing that exploration. Because while you're doing the exploration, you're not going to be very good at getting resources, right? So for humans, in fact for nonhumans too, the way that evolution solves this is by having mothers and fathers and elder parents and babysitters and - my personal favorite, grandmothers - who are all involved in taking care of babies. But the way that they do it is by providing this kind of safe environment in which children or for that matter, adults, can try lots and lots of different possibilities without having to worry too much about the value of those things in the short run. And in a way like what you would like your funding agencies to do, like NSF, would be to do the same thing for scientists or what you want your R&D division to do if you're if you're running a company, is to do the same thing for your engineers, is to say, not here's what you have to do by next week, but to say, some sub category of people, "we're going to just give you some resources, let you be, not insist that you report out to us at every five minutes and then just explore what the space is." But then of course you need the other side, the people who can say, "all right now we've got this good idea, let's see how we can actually implement it." So I think a very important part is actually releasing some of the pressure of actually getting immediate results, is a way that you could encourage this kind of exploration and creativity. And there's a kind of interesting paradox. Another, you know an obvious piece is play. So we have a project where we're looking at computational models of play and play is as an interesting and somewhat paradoxical activity. It's something that we associate with children and that actually young animals in general are playing more than adults. But the very definition of play is that it's not work right? The very definition of play is that it doesn't lead to particular outcomes. But of course what we believe and what we have good evidence for is that by playing early in the short run, not trying to accomplish something means that you're more likely to accomplish things effectively in the long run. And I think it's quite tricky and interesting about how do you, you know, evolution set that up by just having kids who run around and play and aren't doing anything terribly useful at least until they're six or seven years old. And having one of my jokes about this is that under five, children have one utility function, which they're incredibly good at maximizing, as the economists say, and that's be as cute as you possibly can be. And all you have to do is just be cute and the parents and the other investors around you will take care of, you will look after you. And that means that you have the option of exploring. So I think those are all important kind of releasing ourselves from the requirements of outcomes is a good way to allow this kind of exploration. And another thing is just being in novel situations. So one of the reasons why kids explore and play so much is that they don't know as much as we do, right? And there's something to be said for, you know, this is what the Zen masters called Beginner's Mind being in a state where you don't already know a lot, enables you to explore more widely and I think often what you can see in science or in adult creative endeavors is that actually taking on a new problem or a whole new area where your expertise is not going to work, or bringing someone into your team because a lot of times we solve problems in a social way as humans. I was just here reading a fascinating study, looking at scientific papers and it turns out that papers that are authored by people in the same university but in different departments, are the ones that are most successful and most cited. So getting someone who's close enough so that you can really talk to them and interact with them but has a really different kind of expertise and background than you do, seems to be a really good strategy for creative solutions to things. And again, kids kind of get that for free because they are born not knowing as much, but we can try to um you can try to sort of recreate that situation of blessed ignorance in an adult.

And that reminds me of a book whose whole topic was about that interdisciplinary benefit. It's called *Range*. I forget the author for the moment. You had me quite wistful of the description of essentially smoothing the way for the scientists, letting them play or letting them get things done without the stress. I found myself thinking, I wonder how much those babies would get done if they had to answer an email every three minutes?

No, that's exactly right. And that's why if you just sit and watch a two-year-old or three-year-old, they're incredibly busy, like they're doing things all the time, but every single thing they do is about figuring things out. It's not about obligations, or things that they're supposed to do, which is part of why they're kind of frustrating. They can be pretty frustrating to the rest of us.

You were talking about early days of computing a few minutes ago and you reminded me of something, the last guest on my show, George Dyson, who is a historian of computing observed to me. He said - because we were talking about the geniuses, the giants of early computing, people like Norbert Wiener, and John von Neumann, Alan Turing - and I asked him to characterize them and he said that they were like adults when they were children and childlike when they were adults. That was the striking thing when they were adults and that reminds me of what you're saying here or what you're saying reminds me of that, that perhaps there are level of prowess that perhaps their competence was attributable or the creativity was attributable to that playing and exploration.

No, I think that's right and I think there's lots of reasons to believe that that kind of developmental unfolding is exactly one of the things that makes us more or less creative. I've just been writing something that's fascinating news. A [aper that came out looking at what are called adverse childhood experiences. So, children who are growing up with bad things in their environment, poverty violence, neglect. Which is 20% of American children are growing up with these kinds of adverse childhood experiences. Well, you might say, well, what effect does that have on development? Well, it turns out that what it seems to do is actually accelerate this shift from being a child to being an adult. So the kids who are in adverse circumstances are basically growing up too quickly and you can see this in their brains. You can see this even in when they get their adult teeth. So the effect seems to be that, and again, this is echoing this point about you need a mom, you need a nurturing environment, to be able to do this kind of exploration and creativity. You could think about some of what I think happens with the scientists is that, there's sort of the converse that they've managed to extend that period of early creativity even on into adulthood. So when you're supporting child welfare, you know, when you're making sure that you try to get rid of child poverty, and aside from the fact that that's obviously just a good thing to do, the other effect that that's having is enabling a much wider range of exploration and creativity than you would have otherwise. Or at least that's a hypothesis.

So maybe there's some connection here with the stereotype of the absent-minded scientist.

I have to say, you know, my first book was called *The Scientist in the Crib*. And one of my slogans was always that it isn't that and I actually think this is true. It's not that children are little scientists, it's that scientists are big children. That scientists are people in the adult world who we've allowed to continue to exercise these capacities for exploration and creativity. And I've given this talk at a lot of high-powered like Lawrence Berkeley Lab and Fermilab. And every time I say that the physicists in the audience all sort of applaud and nod. They recognize that they recognize that about themselves.

They know who you're talking about. It reminds me of a there was a story about Norbert Wiener who has lots of these stories, but he was walking around the campus one day and stopped to talk to to someone. And then it was typically overwhelming conversation. But at the end of it, he asked, "When we met which direction was I coming from?" And they said, "Well, you're coming from building eight, I think." And Wiener said, "Oh, good. That means I've already had lunch."

I mean, to get back to what adults could do that are that that makes them more creative. I think there's some interesting examples of things like meditation. Meditation this is like, he's trying to find his way across the campus. Meditation is an interesting example where by not doing anything and particularly by not acting, not moving, not just sitting in one place, not thinking in the way that we usually think about thinking at least not planning. People seem to be able to both calm down but also think of more different options. Shake themselves out of narrow ruts that they find themselves in, show more sort of general plasticity than they would otherwise. And I think again this is this trade-off. Now, of course, the thing about being an adult scientist and this would be true about von Neumann and Turing and all the rest, is that you have to get funding for your research and you have to go out into the world. And part of the difficulty of being a scientist is that you have to shift back and forth from This is just going to be the great playful here's a million hypotheses to All right, and now I need this piece of equipment and I need this number of graduate students and we have to decide to do this experiment instead of another. So you're shifting back and forth all the time from one mode to the other.

And some of the most successful projects seem to be the ones where they delegated that to someone else like General Groves, Leslie Groves on the Manhattan Project was the one who took care of the funding so that the scientists would get on with their work and he described it as some sort of cat herding experiment, but it worked. Do you have any words of wisdom, any suggestions for AI researchers who want to create general artificial intelligence?

Well, as I say, look to children, and think about caregiving. So one of the things that I've argued is that if you think about something like the Alignment Problem, which is something that I'm sure Stuart was talking about when he was he was on the program. So the alignment problem is how is it that you managed to design and if you had an AGI how would you get it to have the right kind of values? And actually that’s one of the things that caregivers do too, with children. So one of the things that we have to do as caregivers is figure out, how do we have another, because we have this cultural revolution in each generation is functioning a bit differently in a different environment; how do we allow a new generation to develop values that are going to be well suited for its environment and are going to be beneficent instead of instead of malign. That’s the big scary problem for AI but it’s a problem that we all face every time we have an adolescent child you said you’ve got a 12 year old, so you may not be quite there yet but you will be soon. How do you allow another generation to develop a set of goals and values that are different from the ones that you would have thought of beforehand but are also good ones rather than not good ones. And I think if we ever have and that’s what being a caregiver is all about, that’s what being a parent is all about, that’s what being a teacher is all about. That’s what being a therapist is all about, that’s what being a nurse is all about. There’re all these capacities that humans have for essentially taking care of other intelligent forms and allowing them to be autonomous and to figure out what it is that they want to do and we completely neglect all of that, right? So we ignore the mothers and the grandmothers and the nurses and kindergarten teachers, even though they’re the ones who actually have the skills to solve that sort of problem.

So AI researchers: have more children.

Well, look spend more time with children and spend more time taking care of children for sure.

I mean I just remember that so many experiences that and of course there’s still ongoing, but when they were even younger and thinking to myself three months ago, I was changing this thing’s diapers and now it’s arguing with me, just all these changes. Well, it’s been a fascinating all too short time. Alison, how should people who want to find out more about what you’re doing do that, and what’s coming up next?

So, if you look at my website, which is very simple, it’s AlisonGopnik.com. I have all of my research papers and the work that’s going on in my lab is there as well as the sort of more popular essays and columns and so forth that I write. So you can find either one of those. And in particular, I did a piece about the implications for AI of development that’s right there on n the website. And we’ll see. I’m very interested in this idea, I think an idea that’s very much in vogue now in the air, is this idea of causality might be one of the things that really help solve the problem of AI. That having an AI that could understand causality, understand causal relationships, represent the world in causal terms might be exactly the kind of waystation between the kind of old fashioned symbolic AI on the one hand and the kind of deep learning neural net AI on the other. So having some kind of hybrid system that could take data of the sort that a neural net deals with, but then could understand it in causal terms I think would be very interesting. And that’s part of what we’re looking at now.

I was going to say causality sounds very much like a good old fashioned AI kind of problem: A causes B because of C, therefore A causes C. Going further than that level of predicate logic?

So if you look at something like the kind of work that Judea Pearl, did 20 years ago, we have a lot of really interesting formal characterizations of causality and causal inference and how and it’s a nice example because it’s connected to correlation, right? You’re it’s connected to statistical relationships of the sort that are well captured in a lot of current chaos. So, I think that could be a very promising area of research. And of course, since we know that children are learning a lot about causality, that’s a nice example. Another thing that we’ve done is this kind of methodological thing of actually giving children - it’s interesting that children nowadays with touch screens and so forth, are quite happy to do this - give them information about a system on say an iPad and then see how they explore that system. And then we can give exactly the same data to an agent and see what different kinds of agents do. So how does adding a curiosity bonus to an agent change what they do? It doesn’t make it more make them more like children. And we have a number of different simulated environments, a sort of simulated desk with a bunch of objects you can play on a simulated maze, a simulated machine that lights up and then we can record what the children do. This is part of the point about AI helping us methodologically. So we can actually, it would be hard to just do it with watching a child playing with toys. There’re so many different things they do. But this way we can record exactly what they do in this environment and then we can see whether an agent could do the same kind of thing. So that’s a really interesting exciting frontier.

Well if you’ve got something that involves using an iPad to interact with something and you want a couple of experimental subjects, let me know. Alison Gopnik, it has been wonderful, fascinating; when AI becomes intelligent, self aware, conscious, and wakes up and says, “Where am I?” I’m going to send them to you. I think that would be the best place to start. Thank you very much.

Yeah. Happy to be a recommendation I have is Ted Chiang, the science fiction writer has a beautiful novella called “The Life Cycle of Software Objects,” which is exactly about that, exactly about what would happen if you were raising an AI and all the complexities and that will give you a better summary of the ideas than anything that you can get from me I think.

Wonderful. Alright. Thank you very much. Thank you.

Thank you for having me.

That’s the end of the interview. You can find a link to Alison’s book, *The Philosophical Baby: What Children's Minds Tell Us About Truth, Love and the Meaning of Life* in the transcript. Go to alisongopnik.com to find out more about her, her books, media appearances, and a picture of her of course holding an adorable toddler.

In today’s news ripped from the headlines about AI, British and Indian firm Wysa secured 5.5 million dollars for developing emotionally intelligent AI, an app that would provide cognitive based therapy dialogues. Their app is aimed at people who aren’t so much clinically ill as people who want to improve their sleep, anxiety or relationships, according to Jo Aggarwal, the founder and CEO. We know that people have been quite willing, under certain circumstances, to confide in a chatbot, in many cases more so than they do in many people, and that this goes back to the sixties and the early chat programs like ELIZA, which Joseph Weizenbaum wrote to emulate a Rogerian psychiatrist. He was so unnerved by how much people wanted to open themselves to it that he pulled its plug. Now here we are with Wysa and several other tools like Woebot from Dr. Alison Darcy, that are proving their worth in engaging people in therapeutic conversations. It may seem to defy the usual narrative that the jobs that are safest from automation are the ones involving the most human communication and empathy, but I don't want any therapists to start feeling nervous about their job security. These programs are filling in a niche that's more like a coach that people can turn to at any moment, not providing deep psychological assistance but acting more like a sounding board, and, critically, being available at any time, including 3 o'clock in the morning, and not just when you can get an appointment. It'll be interesting to see how this field evolves.

Next week, my guest will be futurist Calum Chace, author of *Surviving AI* and *The Economic Singularity*. That's next week on *AI and You*. Until then, remember: no matter how much computers learn how to do, it's how we come together as *humans* that matters.

http://aiandyou.net