AI and You

Transcript

Guest: Alison Gopnik

Episode 96

First Aired: Monday, April 18, 2022

Hello, and welcome to episode 96! My guest today is Alison Gopnik, who is the American professor of psychology and affiliate professor of philosophy at the University of California, 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 writing on psychology and cognitive science has appeared in *Science, Scientific American, The Times Literary Supplement, The New York Review of Books, The New York Times, New Scientist*, and *Slate*, and she writes for the *Mind and Matter* column in the *Wall Street Journal.*

She has given a TED talk, been on *The Colbert Show*, *The Charlie Rose Show*, and *Making Sense with Sam Harris*. And she is on the show today because she knows how babies think. No, you have not turned into a child psychology podcast by mistake, this is still AI and You, and you will find out really quickly what the fascinating connection is with and implications for AI as we get into the interview.

Alison, welcome to the show.

Glad to be here.

So, you talk about babies and some people might be wondering why we're going be talking about babies on this show, but all will be revealed really quickly. And you've done a TED talk about what babies think. Is there a quick answer to that? What are they thinking?

Well, they're thinking about the world, everything about the world around them, the objects that people they're figuring out what's happening in the world around them, especially what's happening in the minds of the other people around them. So that's part of the argument.

And the surprising thing about that to me was that they're doing some things that I didn't think were attributable to that age of human development, as you say, they're doing this emotional intelligence, but it goes further than that, and they're smarter in a way that we seem to forget. Can you describe that?

Not only are babies and young children smarter more and learn more than we ever would have thought before, but in many respects, they're actually more well suited to learning, smarter in some ways, than grownups are. And that's part of the reason why they're important for understanding AI; they seem to be the best learners that we know of on the planet. So, in a very, very short time, they learn these fundamental facts about how people and objects in the world and language work, and that seems to be really what they're designed for in those in those early years and the view that people had, even when I started out 30 years ago, that children were blank slates, that they lived in a kind of booming, buzzing confusion. We sometimes hear people

say, oh they're sponges. All that turns out to be wrong. In fact, the analogy that I've used and others have used, is they're more like little scientists, they go out into the world, they have hypotheses and theories about how the world works and then they go out and test and change and revise them in much the same way that scientists do.

And I think we're starting to make some of the connections here with AI. I had always thought of babies' brains as they're forming all these connections, and neuroplasticity is firing at full bore, but it's being used for figuring out much more primitive things like object permanence. Did I get that one right? The idea that if when mommy goes out of the room, she ceases to exist until some age?

Well, that's not so clear. So part of what we've discovered again over the last 30 years, is that in fact even very young children, even very young babies, have more abstract and complicated models, theories of the world, than we might have thought before. So in the object permanence case that you mentioned for example, it turns out to be an interesting, complicated story. If you use some measures children seem to--even very young babies, say three-month-olds—seem to make some predictions about what objects will do out in the world. But they have trouble making other kinds of predictions that would be obvious for grown-ups. And there's a lot of change and learning about objects that's going on in those in those first 12 months, say. And if you actually watch a baby, you can see that they do a lot of things like pick up objects and look at them from one angle and then another, or put one object underneath another and see what happens. And we think that those are all kind of experiments that let you understand and learn things about the world around you. But children are very far from being restricted to their immediate perceptual or sensory motor experience, which is what people thought even 30 years ago, they already have pretty abstract representations of the world around them and then they can change and revise those representations based on the information they see from the world around them.

And this is extraordinary stuff to figure out from something that has a limited ability to express itself that doesn't include language. How do you infer those sorts of things from a human that doesn't have verbal communication.?

Yes, so, of course, that's been the great challenge and the great scientific achievement of the last 30 years has been, you know, even when babies can talk, if you talk to a three-year-old, for example, what you're likely to hear is, you know, a beautiful poem about birthdays and ponies, but not anything that sounds terribly sensible or maybe even is very relevant to what you ask them. So we've had to design techniques often by looking at what children do rather than what they say to figure out what they think. So things like, do they look at one kind of scene more than another scene, what do they reach for? Even with the three-year-olds, what you can do is instead of just asking them free form, what they think, you can give them a choice between two options and see if they pick this option or the other option. Or in the case of the work that we've done about children's causal learning and causal inference, something very relevant to contemporary AI, you can give them a little machine that does things and see if they can figure out how the machine works and what to do to make the machine work. So those are all examples of things

that we can do that let babies and young children tell us just how smart, just how smart they are without having to use language, for example.

Sounds like it has something in common with how we study animal intelligence?

It does except that we face a lot of the same kinds of problems when we're trying to study nonhuman animals. We do have the advantage with babies that we kind of know something about where they're going to end up in a way that is harder when we're dealing with nonhuman animals. And we also know that they're in the same evolutionary niche that adult humans are in. I think there's an interesting other analogy which is that it's very natural for people to, in the case of both nonhuman animals and babies and children, to act as if they're sort of defective versions of us, right? So there's an old idea called the Scala Naturae or the Ladder of Nature. And surprise, surprise, 40-year-old men, human men are at the top and then actually God is at the top in the classical example, it's God, angels, 40 year old human men. And then animals you defined by saying, oh well they can't do these things that we can do. And the same was true for a long time in developmental psychology where people would say, oh well, children can't do things like defer gratification or long-term planning, or certain kinds of representations. But I think increasingly the picture that we have now both for nonhuman animals and for babies and children, is that the right way to think about it is that they have a really different kind of intelligence that's designed for the niches they find themselves in and the problems that they're trying to solve. It's not that they're like a defective version of adult human intelligence. Instead, adult human intelligence involves all intelligence, involves trade offs between different kinds of functions that an intelligence can serve and different creatures, whether they're adult humans or nonhuman animals or babies and young children trade off those functions in different kinds of ways. And what I've argued in particular with children is that children are helping to resolve the explore-exploit trade-off, which of course is one of the classic trade-offs in AI, that trade off between having a capacity to actually act, that's going to be effective that you can do swiftly and immediately and without too much trouble, versus the ability to explore a wide range of, say different solutions or options. And what I've argued is that children and childhood really is about that period of exploration.

Here's the point where some machine learning experts might perk up their ears because you have said that babies are making Bayesian calculations, which is something that we have to wait until at least past high school to teach people. And can you give us an example of how a baby does that?

Again, this is partly about trying to ask babies and children to tell us things in their language instead of our language. So as I mentioned, when we're trying to do these kind of causal learning problems, what we can do is show children say a little machine, we call it the Blicket Detector, it's a little machine that lights up when you put some things on it and not other things. And the problem for the kids is figure out how to make the machine go. So that doesn't mean actually explicitly saying anything about things like probability or models. But with that very simple device, we can present children with patterns of data that would support different kinds of hypotheses, with different probabilities. So for example, we can show them, machine works four out of eight times, or we can show them that the machine works eight out of 12 times, or we can

show them that one of the blocks makes it go, but only conditional on the other block being on the machine. We can show them that the machine has a structure where each individual block makes it go or not, or the machine has a structure where you need multiple, you need combinations that sort of conjunctive structure to make it go. So even with these very simple devices with other little machines, we have we have a machine that either works with a causal chain where A causes B which causes C or common effect. So A causes B and causes C. So we can use these very simple machines to illustrate lots of different kinds of hypotheses about how the world works. And then we can give children data that suggests something about the probability of those different hypotheses. And then what we can do is just ask the children to make the machine go and see what their choice is about the likeliest hypothesis. And what we find again and again, this is now over 20 years of work in my lab, but in lots of others as well is if you present the children with two hypotheses and a pattern of data, that from a Bayesian perspective, should make one of those hypotheses more probable than the other - kind of classical Bayesian inference - the children will choose the hypothesis that has the highest posterior, the one that has the highest probability. Given that data and children are doing that across a very wide range of different kinds of contexts and domains. Now of course, if you asked them what's the conditional independence of Blicket A on Blicket B they would be in fact, I think it's kind of an advantage because when we ask even grown-ups, those kinds of questions, they look pretty dumb because nobody is very good at being explicit about probabilities. That's the whole point of saying the work that someone like Danny Kahneman and Amos Tversky did. But fortunately we don't even try doing that with kids because we know that kids won't understand those kind of that kind of language. So instead we do it in this active way and we can show that children are remarkably intelligent and often doing something that looks like a kind of ideal Bayesian inference.

And still it seems that they are doing something that we can't I mean adults do not have even intuitive understanding of Bayesian probability in many cases, if you look at, for instance, what is it called, the Monty Hall problem with the three doors and if you open one, there's a goat behind it and there's a car behind one of the others. Should you switch or not? Most people get that wrong. And that's a Bayesian problem. When you talk about the babies experimenting a lot and that shows in the in the videos, is this where we are learning our pattern of predicting our behavior continually?

Well, I think one of the things we've shown is we've shown that children are able to solve these problems. And by the way, you know, again, even adults often can solve the problems better if they're in this kind of concrete situation where they're getting data one at a time rather than when you give them a kind of more abstract question about likelihoods or probabilities. But we have shown that children are better than adults when the solution to a problem is unlikely or unusual. And we've shown this now in a number of different kinds of settings and we've also shown that children are unlike adults in that they are more driven to get data than adults are. So adults tend to - sensibly enough - sort of have the attitude, "What I know about the world is sufficient. The main problem I have to face is, given what I know, how can I go out and act on the world effectively?" And children on the contrary seem to be more involved in, "what can I do to find out more, what can I do to get more information?" And what that means is that in certain

kinds of problems the children, because they're so curious and exploratory and because they're willing to think more out of the box and than the adults actually have an advantage can actually solve those problems better than the adults are. So what we're trying to do now, we're, this is part of the DARPA Machine Common Sense project is to try and figure out, okay, we've shown that they can do that; how do they do that? How do they learn those things? Some people have suggested, "Well, they just have that knowledge there to begin with." And that is one possibility. But a lot of the things that they seem to be able to learn, like which of these blocks make this machine go don't seem to be like the sort of things that would be that would be built in. What's allowing them to learn as much as they can? And the picture that one of the things that we've learned is that acronyms are really important in AI so our acronym for our project is MESS which stands for model building exploratory social learning systems. And those are the three things that we think are really the crucial things that allow the children to learn as much as they do. They're building abstract models from data and there going out and doing experiments and exploring - of course when they do it, we call it "getting into everything". And they're learning from the people around them. So they're paying attention to the social cues and the people around them. And lately we've been doing a lot of work looking at that second piece, the exploratory piece, and looking at how just the sort of spontaneous actions that you see in children's play are actually helping them to figure out what's going on around them. And then the really interesting challenge for us is trying to see if we could design algorithms that could do the same kind of thing as the children are doing.

You mentioned something there that is going to ring a lot of bells with anyone trying to work on artificial general intelligence, which is "common sense," because that's often the placeholder for what we don't know how to do in artificial intelligence. And now you say that DARPA is working on a common sense project. What's your involvement with that?

So, this is a project that's designed to be half-and-half AI people and developmental psychologists, with the thought that one of the things that we've discovered, as I mentioned before is that even very young children, by the time they're two or three, already seemed to have a lot of the elements of everyday common sense that we need to be able to function in the world. They understand about objects, they understand about people, they understand about places. And the question is, could we use the model, could you design an AI that was up to the level of, say, a three-year-old's common sense about the world? And you know, spoiler alert is we're very, very, very far from that now. So even if you look at the spectacular advances in AI using things like deep learning, they're very good at taking a very large data set and pulling out the statistics of that data set. They're not so good at actually having a robot that can pick something up and put it in a box, right? Which is doesn't seem like it's such a hard task, this is the old Moravec Paradox where we can get computers to do things that look very hard for humans, like playing chess to playing Go. But we have a terrible time getting a computer that can pick up a block and put it in a box on a regular basis, especially the ones that can do this in a robust way. So that you could, say, move the block a few inches and still or have a slightly larger, slightly smaller block and still get the robot to do it. Again, something that's completely transparent for every 18-month-old. So we want to try to look at these kind of everyday intelligence that we see even in very young children, and young children are interesting to look at because by the time you're talking

about adults, we have generations of culture and history and explicit learning and schooling. So it's really not a very fair comparison to look at us and look at an AI. But if you're looking at children, they're much closer to here's what are sort of foundational cognitive capacities are that would enable us to have the kind of intelligence that we have. And even if you look at something like ImageNet, for example, even when AI systems are good at doing something like categorizing objects, they often don't make the right kinds of generalization. They'll just be stuck at the kind of data that they originally got and it's harder for them to generalize. Whereas, again, every baby, every three-year-old you can give them a task that looks as if it's really new with you know, a toy that they've never seen before. And within the space of a few minutes, they can figure out how it works. So we have this phenomenon of the AI can take very large amounts of data and are not very good at generalizing, we have the kids who seem to use much less data but being much better at robustly generalizing, and the question is, what is it that kids are doing that the AIs aren't. And we think this kind of active learning in the environment might be one of the really important pieces.

You've said that babies are like the R&D department of the human species and that's that that that exploration that experimenting you reminded me of a quote of Steven Pinker's. He says, "When Hamlet says 'what a piece of work is a man, how noble in reason how infinite in faculty in form and moving, how express and admirable,' we should direct our awe not at Shakespeare, or Mozart, or Einstein, or Kareem Abdul Jabbar, but at a four-year-old carrying out a request to put a toy on a shelf."

No, exactly. And the AI example has really made vivid just how much these everyday things that every four-year-old can do, just how complicated they are, how cognitively demanding and taxing they are, how impressive it is that the babies can do as much as they can as quickly as they can.

Has that project got a definition of common sense for success criteria?

Well, for practical purposes, you know, we've chosen a bunch of very specific kinds of abilities like object permanence, being able to find your way through a room, being able to find something that's hidden behind - so, you know, you can do something like you have a simulated kind of playroom and you want the agent to go and get a toy, and the toy is off in a corner behind something else. And the question is, can you figure out how to navigate your way to the toy? So that's a really simple problem. Or can you predict if you see someone else go into that playroom, can you predict where they'll go to look for the toy, or you see someone else and you find out that they want particular they want an apple, let's say; what will they do to try to get to that apple? And again, these things sound very simple and straightforward, they're things that we know that two- and three-year-olds can do, but it turns out to be quite complicated and difficult to figure out how to get an artificial system to have those kinds of capacities.

So when we talk about artificial general intelligence - which we don't have, but everyone wants - one of the ways suggested for creating it is to start with a model of a baby and then teach it. Do you think that's a productive strategy if we could figure out even how to start with a baby? Do you think "baby" is as easy as they think?

Well you know, there's a beautiful quote from Alan Turing from the original paper in which he talked about the Imitation Game where he says, maybe, after he's gone through the whole business of here's the Turing Test and here's what it would mean for a system to be intelligent, he says, "But maybe we're doing this all wrong. Maybe instead of trying to develop a machine that simulates an adult, we should instead try to make one that simulates has the powers of a child." And the reason why Turing says that, and the reason why this has become so much of interest in modern AI, is that if what you're trying to do, which was kind of good old-fashioned AI's project to build in the kind of competence that you see in an adult, that's one approach. The other approach is to try and get a system that doesn't have very much built in, but it's very good at learning from data. And the great advances in the last 10 years or so, the great AI Spring, has really been based on finding better and better ways of getting machines that can learn from data. And if you want to have a system that learns from data, the best example we have is a child, right, that's what children are spending all their time doing, that's what they're designed to do. And in fact, as I say, often they're much better and more motivated to do that than adults are. So I think the general idea is it looks as if the route to AGI is especially for creatures like humans, especially when you want an intelligence that can adapt to new environments that can deal with new situations, is to have a system that learns. And then the question is, is there a system that can we use these hints from looking at the children who are the best learners that we know of in the universe to design the kinds of learning that would be relevant to getting a genuinely broad intelligence.

And if you're familiar with the work of Mark Sagar at Soul machines and Baby X. Do you know ow close that might be getting? I've just seen some documentaries.

He has some lovely pictures of babyesque-looking creatures but they're very far in terms of their capacities from what we can see the kids around us doing. And in fact that's been the challenge I think in this whole project has been, let me give you an example. We're just doing a project one of these causal inference projects about figuring out how one of these machines works right? You've got a bunch of blocks, you have to figure out which ones make the machine go and basically how the machine works. So we've done this experiment now and what we do is just give the kids the machine and say, "okay, figure out how it works." And it takes them about 20 trials and they figure it out. If you take a pretty state of the art reinforcement learning agent and put them in exactly the same situation and were using an on screen version of this Blicket detector so we can give them exactly the same, we know they're getting exactly the same day that the kids are getting it's they can solve it but it'll take them 100,000 trials before they can solve it. So that's just a giant gap between what you know very powerful learning mechanisms that we know of currently in AI are doing and what the children are doing. And as I say, what we're trying to figure out is how could we bridge that gap? What are some of the things, what are the representation techniques that the children using that are letting them letting them do that?

And so what I've thought about a lot is that humans, we seem to be framework-building machines, that as we every time we learn something we use it to improve and extend the interior models we have for learning more things. And I just made that up. I don't know if it's real, but it seems like it to me. And that that is the thing that we don't seem to know how to do

in AI at the moment. The reinforcement learning models that you're talking about have to start from essentially knowing nothing about the world whatsoever and get trained on all this data and then that's the only thing they know how to do. Whereas a child can have a one-shot learning of what a cat is and they'll get it in one go. But they already have some experience with furry things and living creatures and all that framework to put it in and then they can put the cat into that and then they extend it in the direction of feline family and things that purr and so forth. Which obviously is very effective for us. But I don't know if anyone's got any idea how to do it in AI. What would you say about what I've just been talking about?

Well, I think the basic tension - which is not just a tension in AI, it's actually a tension that goes back to the very beginnings of philosophy talking about knowledge, back to Plato and Aristotle - I think there's this very basic paradox about how we know about the world. And that paradox is, as you say, we seem to have these very abstract structured representations of the world around us. And it's having those abstract structured representations that enables us to do things like make good generalizations. So if in my head instead of having okay, here's a bunch of pixels that I observed of this fuzzy textured thing, I have a representation like "cat" that lets me generalize and make predictions in a much more general way than I could if all I had was a combination of the pixels. And again, this is not a new insight: going back to Plato people have pointed out that we need those kind of abstract representations, that's what lets us do all the have all the kinds of knowledge we have. On the other hand, it looks as if we get those representations from a bunch of data, a bunch of disturbances of air at our ears and photons hitting our retina that don't have any of those characteristics, that aren't abstract and hierarchical and structured. The data really is pixels. IAll of us are in a world where what we're interacting with out there in the world is a bunch of pixels. So the puzzle going back again to Plato and Aristotle, the way that the that philosophers and psychologists have approached this problem, is either to say, "okay, it just looks as if we learn all this. It must be built-in." Or else to say, it just looks as if we have these abstract structures. Really if you just have enough data, it'll be fine, right, you don't need to have the abstract representations. And I think if you're a developmental psychologist, neither of those seem very satisfying because what we see when we actually look at children is that they have, as I said before, they have amazingly abstract structured hierarchical representations from the time they're very young, probably from the time they're born, and yet they seem to be changing and revising and altering those based on the data they get. So children in a very vivid way have this really fundamental question, which is how can we put together the data that doesn't have all these kinds of characteristics? How can we construct representations? How can we construct categories? How could we construct abstract models from data that doesn't start out that way? And I think that's the really basic foundational problem for AI. And the foundational problem for understanding human intelligence too.

That's the end of the first half of the interview; we've split it into two episodes to make it more digestible for attention span and download times.

This is pivotal research for figuring out artificial general intelligence. If we want an AI that thinks like we do, that solves general problems, and most importantly, has the ability to learn and get better at learning, I don't think we're going to make much of a dent on that until we know a lot more about how

humans do that. I could be wrong, but I don't think so. And what Alison's made me aware of is that there's a lot more going on in the minds of babies than I first thought. As you heard, they're not just undifferentiated blobs of neurons slowly learning the most basic biological processes, but they're actually employing some advanced scientific-method-types of thinking, they just can't language it in the way that would make it clear to everyone what's going on, but they're actually doing better at that kind of thinking than many adults.

We referred to Bayesian calculations, there, by the way, are used heavily in AI, and refer to conditional probability. Bayes Theorem tells you how to calculate the probability of one thing happening is if another particular independent thing has happened. The Monty Hall Problem I referred to is fun to explain—just to be clear, I wasn't saying that babies would be able to figure this one out, but it's a problem that many, if not most adults get wrong. It's named after an old game show host, Monty Hall, who had a show called *Let's Make a Deal,* and one of the games in that, or at least one that we imagine, would be a single contestant faced with a stage containing three doors. So say you're the contestant. Monty tells you, "Behind one of these doors is a new car—you want the car, right? And behind each of the other two is a goat—and I'm assuming you don't want the goat. You don't want to have to take a goat home with you, right? So the game is to pick the door with the car. So choose a door." And of course you have no information to go on so you pick a door at random. But instead of opening that one, Monty—who knows what's behind each door—opens one of the others, and reveals a goat. Then he asks you, "Do you want to stick with your original choice, or switch to the other door that's not opened?" Pause the show for a moment if you need, to decide what you would do. Are your odds better if you stay put, switch, or doesn't it make any difference?

Okay, if you said it doesn't make any difference, that's what most people think—and it is wrong. This can sound crazy—as though the car somehow switched places while you were deciding. But in fact you double your chances of winning if you switch, and here's why. The first door you pick has a one in three chance of having the car behind it, right? Which means the chance of it being behind one of the others is two thirds. When Monty opens one of the others, he knows which one has the car. He's not going to open that one. That's the key piece of information. You know he's picking a goat. So the chance of the car being behind one of the other two doors is still two thirds, but now one of those choices is eliminated. Therefore the chance that it's behind the remaining closed door is two thirds, which is double the odds that it's behind the one you first picked. Therefore you should switch your choice.

In today's news ripped from the headlines about AI, not about talking to babies, but talking to pigs. Researchers from the University of Copenhagen, the ETH Zurich, and the National Research Institute for Agriculture, Food, and Environment (INRAE) in France have used AI to figure out the language of pigs, more or less. No, it's not pig latin. The study is titled, "Classification of pig calls produced from birth to slaughter according to their emotional valence and context of production," and it appears in the Nature family of journals called *Scientific Reports*, and they were able to train an AI to determine whether a pig noise indicated a positive or negative emotion. They found that screams and squeals are associated with negative emotions, whereas barks and grunts – very tempted to do an impression, but think you know what that sounds like - could be either positive or neutral. They did this with a cluster analysis machine learning algorithm called t-distributed Stochastic Neighbors Embedding. So who knows where this is going, but maybe a Dr. Doolittle app for your smartphone isn't that far off in the future.

Next week, we'll conclude the interview with Alison Gopnik, when we'll talk about what babies learn versus what they're born knowing, what we can learn from babies, a connection with prominent scientists, and even a connection with the AI alignment problem. 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