AI and You

Transcript

Guest: Kristóf Kovács

Episode 10

First Aired: Monday, August 24, 2020

Hi! In this episode, we'll be talking with Kristóf Kovács, the supervisory psychologist of International Mensa. Mensa is a worldwide society whose membership requires a score in the upper two percent of the population on an IQ test. I actually passed the test and joined British Mensa when I was 16, which is not a way to be popular in high school, but more about that another time. It's an interesting organization; there's a huge range of interests, personality types, degrees of personal success, but everyone is very engaged in *thinking*. Obviously, it's also very much an organization of people who want to join it in addition to being able to join it, so it's not necessarily representative of everyone who has that level of intelligence. And by general agreement, no one talks about their IQ, they all treat each other as equals. I spent a few years there, and revisited it again not long ago.

Kristof speaks for Mensa on the definition of intelligence. His job is to ensure that they're measuring the right thing in the right way, and he's an active researcher in cognitive psychology and psychometrics. We've spent all this time talking about artificial intelligence and we know what 'artificial' means, but what is 'intelligence'?

But first I have to give a plug for something I'm very excited about. For the last three years I've been giving a continuing studies course that covers the same theme as this podcast: What is AI, Why will it affect you, and How can you and your business survive and thrive through the AI revolution. And this year, for the first time, that course will be online! For fairly obvious reasons. So I had the time changed to morning in my timezone – Pacific time - so that it runs at a time where everyone from Hawaii to Moscow can participate at a reasonable hour. Sorry, all you folk in India, China, Australia, and Japan. We'll work something out for you next time.

This is 10 *hours* of instruction. It takes place over 5 classes, one per week, starting September 9. There's a registration link in the show notes and the transcript. Or you can go to continuingstudies.uvic.ca and search on artificial intelligence, it's the first hit. You'll know you're in the right place when you see University of Victoria at the top. Since it's online I'm hoping they won't cap the signups at the limit they have on the page. I get paid the same no matter how many people are in the course, the reason I want lots of people there is to get more of my message out there and because the more people the more fun and impactful the classes.

What are we going to talk about? A huge variety of things, from the history of AI, to the present issues, to the speculative future; from the people who are influential in the field to the impact of AI on jobs, media, and society. We'll spend a great deal of time explaining AI at a practical level – that doesn't require computer experience – so you get a good idea of just what it can and cannot do now and in the future.

Obviously that's a really broad syllabus. Just like this podcast. We're not going to teach how to program AI, there's no code or math, it's all about… well, AI and You. *Everything* I'm producing is doing that job: this class, my videos, my TEDx talks, my book, and this podcast. Next up there will be a Broadway musical.

My vision – just so you know where I want to take this – is to produce not just a course but an entire department, giving multiple programs of multiple courses, for credit, at undergraduate and graduate

levels, and also with high school and corporate training versions. The idea is not world domination – well, not just world domination - but to create part of what we need to help people understand how to deal with, and leverage, disruption.

https://continuingstudies.uvic.ca/humanities-and-social-sciences/courses/artificial-intelligence-and-you

Alright! On with the interview.

Welcome, everyone. This is going to be a fascinating show because my guest today is Kristóf Kovács, who is the Supervisory Psychologist of Mensa International, which is a worldwide organization, whose only criterion for membership is that the members score in the top 2% of a standardized test of intelligence. Kristof obtained an MA from the University of Szeged, and a PhD from the University of Cambridge. Subsequently, he worked as a postdoctoral fellow at the Budapest University of Technology and Economics and the University of Amsterdam. His main interest is individual differences in cognitive abilities, bridging cognitive psychology, and psychometrics. Currently, he is senior research fellow at, and I'm going to give this my best attempt here, at Eötvös Loránd University. Did I get that right, Kristof?

Yes, you did. Hello, and thank you for having me.

Well, thank you for coming on this show. This is going to be really interesting. How did you get into this field? What led you to be interested in the nature of intelligence?

Already as an undergraduate I was interested in it, so this was the topic of my thesis or something related to this was the topic of my master's thesis, just individual differences in cognitive psychology. Individual differences are pretty much sort of the abandoned child of cognitive psychology. Really, it developed in its own ways from the mainstream study of cognition. So the interesting thing in psychology is, so on the one hand, you have a field that studies cognition, and another one that studies individual differences in cognition, and they are very different in so many ways. The initial difference is research is very applied and I was interested in intelligence from a cognitive psychology perspective. So, not psychometrics primarily, even though eventually, for my research, I had to get my hands dirty with some psychometrics because if you study individual differences, it's unavoidable. But mostly I have a basic research interest in individual differences rather than measurement interests. So, I think it's a very interesting area but that's not as popular in cognitive psychology as many other areas and as much as it should be, in my opinion.

Thank you. And now the relevance of this to artificial intelligence is that well, if we look at other things that are artificial, artificial turf, artificial sweetener, artificial insemination, the antecedents of those terms are precisely defined. We know very well what makes the artificial versions of them artificial. But when we talk about artificial intelligence, we have a great deal of difficulty defining the term, partly because we don't know what intelligence is in humans, or at least most people don't have a good understanding of that, and that's why we're talking with you here. How should we understand intelligence? Most of us fall back on the definition of well, “Intelligence is what's measured by IQ tests”, which is great if you're not the person who has to write the IQ test, but I think that's something that you do, right? You have to know how to do that. So, help us out here.

Okay, the many people include most researchers, because there's no universal agreement on a definition, and it really depends on one's approach. If you look at a battery of IQ tests- okay just

one step back. IQ is measured in group settings as well as individual settings. What Mensa does is it usually measures in group settings and then you get just one or two kinds of items in terms of domain, but individual IQ testing happens with test batteries that have like 10, 12 different subtests. And if you look at them, your impression surely will be that they measure different things. I mean, a given subtest requires vocabulary, another one, non-verbal inductive reasoning, a third one, something spatial, like rotating objects and telling which one's identical and which one's mirrored. Then there are tests of general knowledge. There are tests of pure speed which means that they are so easy that given enough time practically anyone could get a perfect score, but since there's a strong time limit, people have to go as fast as they can. There are, of course, tests of memory when you are told, let's say, digits and you have to repeat them in the original order and reverse order. Probably most people, if you just give it to a number of laymen and ask them if they measured the same thing, my bet would be that most would say no. I mean, these are different things. It's one of the oldest debates in psychology, whether to focus on a general intelligence, I mean, if there is a general intelligence that sort of permeates all cognitive activity, or there are a set of specific abilities. Now, if you look at these particular test batteries, what happens is that they administer these subtests, and then it's possible to calculate some kind of global score, which is usually called some kind of IQ but not necessarily. The most widely used individual test, it's called Full Scale IQ but another one which is also widely used, and in my opinion, one of the best ability tests. I don't know if I'm allowed to say the name of tests or that's advertising. Anyway, it's the Woodcock-Johnson Tests of Cognitive Abilities. It's arguably one of the best tests of intelligence and it doesn't have intelligence in its name. It talks about cognitive abilities in plural and the index is called general ability index. It doesn't have intelligence or IQ in it, whereas the Wechsler test, which is probably the most widely used test, has a global index for Full Scale IQ and it's the Wechsler Scale of Intelligence, it's singular. So these are different approaches that manifest themselves in actual test construction. One approach is that we have these different subtests, but frankly, all of them just measure the same thing. So they measure the same thing with different content and to different extent. But they measure the same thing and that's called the general factor, which is something I'll get into. So you could think of thermometers as an analogy. You can measure body temperature in a number of different ways. You can have infrared ones, ones that need contact, ones you put in your mouth, ones you put them under the armpit, ones you get next to your forehead, and so on and so forth, and they might give different results. People with small children are very well aware that they annoyingly often do give different results. But nobody would really think that they measure different things just because they give different results. And there's pretty much a consensus I think, that body temperature exists without measurement. It would exist if no one measured it and these are just imperfect measurement devices that give disagreeing results because of error terms. Now, the approach which says that all of these tests measure the same general intelligence basically are three tests as analogous to thermometers. Whereas another approach focuses on specific abilities and says that you can get a weighted average such as IQs and any kind of global index, but they will be just that, a weighted average. So, it's a summary statistic of a few very different things. So then it's not that everything measures the same thing, just to a different extent but these things- and you can still constitute an index. There are a lot of index variables used in life. Like for instance, the Global Competitiveness Index is composed of a dozen different things that are still correlated, just like in IQ, but they're different. So that gets back to our definition, people as researchers have different perceptions and the big question is general intelligence. Now, if you administered these 10, 12 tests to a population, you will find that results will correlate. So if you correlate any test with any other test, the correlation will always be

positive. That's called in the literature, the positive manifold, which is technically just a correlation matrix with positive entries. If you use a statistical method called factor analysis, which has been thoroughly employed by intelligence researchers, it will give a general factor, so it will give a single factor that explains about half of the total variance. The debate revolves around whether the *g*, which is a statistical construct, where the *g* actually represents something psychological and if yes, then what? Or if it's not necessarily an artifact, but still just a statistical construction and then it's not that best to reflect *g*. So the debate revolves around whether *g* is the common cause of the correlations between tests or the common consequence of those correlations. Now, I'm not trying to be difficult and avoid the question, but you see that this is the fundamental question in which issues differ, and with different approaches to this question lead to different approaches to the definition of intelligence.

It is a difficult question, I think, for many people to grapple with. So you're saying that you can look at different kinds of tests and you can statistically determine whether they are measuring the same thing. It's the nature of that thing that's still up for debate. In say, the Self Help section of bookstores, we can find things talking about different kinds of intelligence that- for instance, emotional intelligence is a thing. And people talk about seven different kinds of intelligences in some models and then physical intelligence. If I look at the capabilities my older daughter has, she's got horticultural intelligence. She can make anything grow much better than me. My wife has culinary intelligence. There could be a nutritional intelligence of knowing what the best things are to eat intuitively, and these could be considered, perhaps domain-specific intelligences, but there's still some underlying factor that seems to be different. Like we should separate intelligence from knowledge, right? If we think of the brain being analogous to a computer, then somehow it seems as though what we know is the RAM and intelligence is the CPU, the thing that processes it and that could have different capabilities independent of what we know. Is that true? How much does intelligence depend upon memorization?

Okay, so these, I think are two questions. The memorization, again, they also depend on if you are a believer in general intelligence or not. So, you said there are seven different intelligences, you probably meant Gardner's intelligences…

Yes

…But actually there are models of purely cognitive abilities that presume seven or even more abilities. For instance, one of the most accepted models is the CHC model, which CHC stands for three people, Cattell, Horn, and Carroll. The CHC model actually proposes seven main different abilities all of which are cognitive. And what the CFC model is built on, or is the successor of, is the so-called fluid crystallized model, which differentiates between fluid and crystallized intelligence and it seems similar basically adds more abilities to this distinction, so, not just for fluid and crystallized. The difference between fluid and crystallized intelligence is that fluid intelligence reflects the ability to solve novel problems. When we have to solve, we cannot rely on the acquired skills or knowledge. Crystallized intelligence, on the other hand, is the application of our acquired skills and knowledge and a typical test of crystallized intelligence is vocabulary. Now, obviously, there's more knowledge component in crystallized intelligence because that's pretty much defined as crystallized knowledge and knowledge of one's own culture pretty much, whereas fluid intelligence is novel problem-solving. So if you think that this is a meaningful distinction, which I personally do, then the answer is that it depends not much in fluid intelligence,

but a lot in crystallized intelligence. As for the domain specificity thing that you mentioned, I think it's very, very important. And I think it relates to our little the problem of artificial intelligence because in the golden age of artificial intelligence, like in the 1950s 60s, when people believed that a computer will be the chess champion in like 10,15 years, which didn't happen for another 30, 35, depends if coming from 50 then even more. So then they were hoping for a human-like intelligence. One that is very general so that it actually comprehends different sorts of problems. Whereas what artificial intelligence really managed to create, and this isn't just my opinion, Marvin Minsky, one of the founding fathers of artificial intelligence research, once wrote about this that what in fact happened is that a lot of specialists have been created. So many artificial intelligence agents are like your family that you mentioned. They are experts in agriculture or cooking or things that are also a particular specific field. And the interest in a human-like intelligence that's artificial is still on the table of course so there is interest in this, and the question is on the table. These days, it's very often called AGI, you're very well aware, which is Artificial General Intelligence. But it's a funny evolution, I mean, AI was meant to be AGI 40 years ago. What happened is that AGI really never was created the way it was desired them, and then what was created is a number of very good specialists, and so there's a born interest in that. So as I said, this is similar to the debate about if there is a general intelligence in the first place in humans because if there is not then it's going to be a very tricky enterprise to create an artificial version of it if it doesn't exist. So, you see these problems are interrelated.

Right, and you mentioned culture there, which is a, I think, a very important factor because weren't some of the early intelligence tests quite culturally biased? That they had questions about hypothetical situations that you had to solve and they were predicated on certain cultural assumptions that wouldn't necessarily apply to everyone that might be taking the test. So they would exclude a lot of people, artificially lower their score because they weren't part of the culture that wrote the test.

Yes, that is very much the case. Early IQ test had a lot of cultural component without test constructors giving adequate acknowledgement of this fact and the results were often interpreted as general, innate and very fixed ability, all of which, that I just listed off which are of course controversial and are indeed a matter of debate in the field. And yes, there are many examples where I cited a number of books, particularly ones that criticize IQ tests, and there a number of such books as we know, and they cite examples from early IQ tests. Tests like where one had to interpret situations that are obviously bounded in culture. In Binet's first test there's this famous figure where there are pairs of female faces and one has to decide which one's prettier, and so on. But this just to make clear, it's not just the early IQ testers, there is no such thing as a completely culture-free test. So there was something people hoped for, and there still are, to be honest, but I think the arguments are pretty strong there is no such thing as completely culture free. So even non-verbal inductive tests, like the Raven's, if you're familiar, where you have to solve visual matrices and there's no learned material. It measures novel problem-solving in western countries. Still, that's dependent in many ways on schooling and if you bring it to some places in the world where the people have had absolutely no formal schooling, then it won't measure the same thing. There are psychometric techniques now to measure one. Well, mostly what you describe is called test bias if it's at the test level or item bias if it's at the item level, and there are statistical techniques to detect either. So if the probability to solve a given item correctly depends not only on the latent construct that's being measured but also on one's group membership, then the test is biased. I try to explain it because it's a complicated thing. So just because there are differences between groups,

it doesn't necessarily mean that the test is biased because it can be the case that there are differences in the latent construct that is being measured by the test. But if someone in one group, with a given level of ability - a given standing on the latent construct - has a different probability of getting the item right, than someone in the other group with the same exact level of ability, then the item is biased because then the probability of getting the item right is not only and exclusively conditional on ability. Yes. So I don't know if that answered the question.

I think it starts to open up a really interesting question. In artificial intelligence, people try to define that in terms of the ability of some object, some entity, some agent to proceed towards an objective, to achieve a task in a world, but it's the definition of what that world is that I think is critical here. And that also is what gets at the difference between cultural interpretations. You could take the smartest person in the world, Marilyn vos Savant or whoever it might be these days, and, like in the movie Crocodile Dundee, you put her down in the bush in the Outback of Australia or New Guinea, and one of the locals there will be far smarter than her at surviving because she's from our culture and urban culture, and she's going to be smarter than any of us at dealing with a new situation, but not as smart as someone that's got that capability of dealing with a really different world. So, it seems that to evaluate intelligence, we have to define the world in which that intelligence operates. Does this make any sense?

Yes, it makes perfect sense actually. And what you touched on really, implicitly is the concept of national IQ and these maps of IQ scores in the world with the countries having different average IQs and if you've seen these, these are loved generally by the media interval, popular news sites but it's also very controversial. Yes, my supervisor in Cambridge, Nick Mackintosh once wrote a very strong review of one of these books on national IQ in which he criticized actually what you just said. Like you can bring an IQ test to the Outback like a Crocodile Dundee situation, left alone to the indigenous peoples living there and find a low IQ score. And they will predictively score low but it doesn't mean they are not intelligent arguably. So my supervisor's example in this book review was Bushmen, the San Bushmen, which according to one of these books, on national IQ, are supposed to have an IQ of 54 because that's how they perform on IQ tests. Now, an IQ of 54 pretty much equals the mental age of an eight-year-old European or American child. It is not very likely that you'd drop an average American or Canadian or European eight year old in the middle of the Kalahari Desert and they will survive, whereas San Bushmen clearly do and they have for tens of thousands of years. So yes, this is a tricky thing. You can bring westernized tests- and it leads back to what I said is some people think that there are completely culture-free tests and it is possible to test the exact same abilities. Just bring an IQ test to the Kalahari Desert and give it to Bushmen, and the test will measure the same construct there. I don't think so. I'm not convinced that that is the case, so I agree with what you just said.

So it seems that to have a useful definition of intelligence, we have to have a useful definition of the environment that it's operating in. If I define that environment to be the game of Go, which is a pure abstraction that we just represent with black and white stones on a board, then AlphaGo Zero is the smartest entity on the planet. But that's not useful to us because we operate in a world that's many times bigger than the game of Go and it's not that interesting to most of us. So I think we have a good idea of what we would like artificial intelligence to do if it was generally intelligent - sweep the house, do all kinds of things that involve an understanding of our local environment. And that's the big question as you say, general intelligence is "How do we get

something to have that general understanding of the world?" But an artificial intelligence that started evolving, according to say, some abstract environment, say that it grows out of understanding mathematics, theoretical physics, chemistry, we start piling Wikipedia in there and it becomes as smart as Watson but it never grew up in the world. It never moved around. It didn't have that kind of background. How should we start evaluating whether something we're confronted with that claims perhaps that it has general intelligence really does?

I think what one has to realize is that the concept of measurement of intelligence is very different for an artificial agent and for humans because if you think about tests, most, until very recently, practically all tests of artificial intelligence, give sort of dichotomous results. So, if you think about the Turing Test, you can either pass the Turing test or you fail the Turing Test. Human IQ testing is very different. So the purpose of human IQ testing is to put the put on a scale to get a quantitative evaluation of their cognitive abilities. In humans, it doesn't happen where people take an IQ test then "So what was the result?" "Well, you're intelligent, that's the result." or "You are not intelligent". You pass the IQ test, or you fail the IQ test, you don't say things like that. On the other hand, when it comes to machines, until very recently, every approach was like the Turing test - you fail or not fail. Recently there have been attempts to quantify artificial intelligence, so get an artificial version of an IQ of some sort. So a version of the IQ for artificial agents. I think that is a very fundamental difference in terms of measurement.

So – you guessed it – we'll be concluding the interview in the next episode so that we don't exceed certain time and size thresholds for podcast episodes. (It's always time and space we're short on, isn't it?)

We mentioned the Turing Test there, and that's something that will come up a lot in this podcast, but the quick explanation is that it was proposed by Alan Turing as the way of deciding whether a computer is thinking like a human, by having a conversation with one without knowing whether you might be talking to a human, and seeing whether you thought it was in fact human.

One of our quick looks at the latest headlines about AI here. Could AI be smart enough to discover or invent new theories about the nature of the universe? One already is. Tailin Wu and Max Tegmark at MIT have developed an AI that copies some of the methods of Galileo and others. It's called the AI Physicist and is capable of teasing out several laws of physics – now here's the catch – it does that in simulated worlds that have been deliberately constructed to model the complexity of our universe.

AI systems often produce overly complex models to describe the data they've been trained on. What's special about this system is that it's been trained to prefer simpler theories over more complex ones, which you may recognize as the Occam's Razor approach that is one of scientists' tools for constructing theories.

Another approach physicists use is to look for ways to unify theories. Look at Einstein. Physicists have serious Einstein envy because he came up with E = mc2 and not only is that just about the simplest equation in physics but it also unifies matter and energy, which is most of what we deal with. Every physicist dreams of finding another equation that simple that has that kind of meaning.

And finally, another principle that scientists apply to creating theories is lifelong learning: the idea that if a particular approach has worked in the past, it might work in the future. So the AI Physicist remembers previous solutions and tries them on future problems.

So Wu and Tegmark created 40 different simulated worlds that had different laws of physics in them from what we have in the real world, and let the AI Physicist see if it could figure them out.

And it did, and it did it a lot faster and better than a conventional neural network. In fact its error rate was about a billionth of what one of those was. So this points the way towards AI helping scientists develop new and improved models of our universe. Again, I have to point out that we're not by any means at a general artificial intelligence, but we really are showing just how much narrow AIs are able to do.

In the next episode of AI and You, Kristof and I will talk more about the nature of intelligence, whether you can increase your own, whether the average level of intelligence in the population is changing, and just what an IQ score means. We'll also talk about what Mensa is like, and Kristof's research into intelligence and genetics.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