

# Cognitive Psychology and the Limits of AI: An Interview with Dr. Joe McDonald

<https://mindmatters.ai/podcast/ep327>

Robert J. Marks:

Greetings, and welcome to Mind Matters News. I'm your Cognitive Host, Robert J. Marks. Today we're going to talk about cognitive psychology and artificial intelligence. What is cognitive psychology? Well, this is what I understand. Earlier, traditional psychological studies focused on what you see, on what humans do. Cognitive psychology explores inside the person. It concentrates on human thought and decision-making. Cognitive psychology doesn't just admire the appearance and the performance of the car, it looks under the hood to see what makes it tick. Cognitive psychology investigates how humans think, learn, and process information by examining mental activities like perception, memory, reasoning, and problem-solving. It delves into the intricate workings of the mind using methods to understand how we acquire, store, and use knowledge.

To talk about cognitive psychology and its relationship to AI, our guest today is Dr. Joe McDonald. Joe is the CEO and Principal at Jumpseat Research. It's a human factors and user experience research firm right here in the great state of Texas. He's in Austin. He has 15 years of experience in applied cognitive science and product usability. Dr. McDonald holds a PhD in Human Factors from Georgia Tech, and he is pretty passionate about empowering people through technology and advancing standards such as the human readiness level within the Department of Defense. Joe also has some informed opinions on the psychological impacts of AI-generated language and human thinking. So this should be a fun chat. Joe, welcome.

Joe McDonald:

Well, Bob, thank you so much, great to be with you. I always enjoy sitting down over a cup of coffee to chat cognition, and it's so great to do it here on the show, so thanks for having me.

Robert J. Marks:

It is good. Okay, I've got my coffee here, but it's iced coffee.

Joe McDonald:

Right. Same.

Robert J. Marks:

Is that okay? Oh, you have iced coffee or just regular coffee?

Joe McDonald:

I do. Yeah.

Robert J. Marks:

Oh, man.

Joe McDonald:

Iced today.

Robert J. Marks:

Okay. I've got to ask you about your company named Jumpseat. I read that and I thought, "What's a jump seat? What's a jump seat? Why is your business called Jumpseat?"

Joe McDonald:

Yeah, yeah. Well, so the jump seat actually goes back to the very foundation of what we call human factors research that you mentioned earlier. So applying cognition to things like cockpit design in an aircraft, which is where that name came from. So in World War II, the US was concerned with fighter pilots and how they can react very quickly in these dogfights by things like touch and feel and look and sound within the cockpit. And so, that jump seat was a place where researchers could sit and observe what happened in that cockpit and make informed design decisions about how to do that quickly and more effectively based on the way people think about the world around them.

Robert J. Marks:

Okay. Is the jump seat the thing you touch and it throws you out of the aircraft?

Joe McDonald:

No, that would be the ejection seat. You don't want to be sitting in that one.

Robert J. Marks:

Okay.

Joe McDonald:

Jump seats are the ones you may see up in the front right behind the cabin door, where you see maybe pilots sitting on that to catch their next flight or flight attendants may sit there while they're waiting to serve you.

Robert J. Marks:

Okay. Does it have anything to do with jumping out of the airplane?

Joe McDonald:

No. No.

Robert J. Marks:

No.

Joe McDonald:

They used to, for folks that were jumping out of airplanes into say theater in battle, they would call those jump seats as well. But I'm referring to more the commercial side.

Robert J. Marks:

Okay. And what's the relationship between your company, which is Jumpseat Research, and the jump seat in the aircraft? What's the connection there?

Joe McDonald:

Yeah, absolutely. So the original researchers who looked at how pilots fly planes and how the crews interact with everyone on board would sit in those jump seats and observe what was going on. So that gave them a chance to get a firsthand perspective on how we can use things like information processing perception, taking notes on the ways that they navigate that plane and communicate with air traffic control, and then designing better systems around that. So it's a throwback to, we love being out in the field and sitting where people interact with technology and using cognitive psychology principles to really inform better design.

Robert J. Marks:

Okay. Okay. Now I gave a stab at defining cognitive psychology. You're the expert with the PhD, and I think you could probably do a better job. What is cognitive psychology? And my understanding is that cognitive psychology has changed over time. Can you give us a brief history of cognitive psychology and what we know about it today?

Joe McDonald:

Yeah, yeah. I think you're selling yourself short. It was a great description. But I think what most people kind of think about psychology, they really kind of picture emotions or introspection, perhaps a couch and a notepad and somebody asking how that makes you feel. But cognitive psychology, however, takes more of a technical term, so investigating how we process information and use that to make our way through the world. For really over 70 years now, cognitive psychologists have worked to unravel those mysteries of kind of the black box, as we call it, between your ears, so the processes that transform inputs from the world that we see and hear and smell into outputs of thoughts and behaviors. And so, we've done this using scientific methods.

So as your listeners know, the scientific methods starts with an observation of some phenomenon in the world. For example, you may, and this happened to me recently, walk into a busy room at a conference looking for a colleague. And they could be right in front of you waving their hand to you and you completely miss them and walk right by. As scientists, we may pose a testable question here called the hypothesis about that. Was the color of the friend's shirt blending in with the background? Did the noise in the room drown them out? Or was it that distracting vendor blimp that was flying around inside the room that drew your eyes away? Well, we'd have to design an experiment that looks at this thing called attention, and strip away each of these elements in the environment to see how each of them did contribute to this mistake of walking right by your poor friend.

So this phenomenon in particular is called inattention blindness, and although the example I gave is relatively harmless, you can imagine the safety concerns of doing this on the highway. When you're focusing your attention on cars on the road, it's actually more difficult to notice motorcycles because we're scanning for things that look like cars in the size, shape, motion, driving patterns of these large four-wheeled vehicles. So this is one major reason why motorcycle accidents happen more frequently percentage-wise than car-on-car accidents at base rates considered. So these mental processes and errors that cognitive psychologists call constructs are really persistent across tasks, activities, domains, and they've been observed long before modern technology such as smartphones have existed.

Robert J. Marks:

Okay. Yeah, I want to talk later about some of different psychological tests and how people are going to respond to them differently than artificial intelligence. I find that just fascinating. Now, I want to talk about behaviorism. Before we talk about something, we probably should define it. What is behaviorism in terms of cognitive psychology?

Joe McDonald:

Yeah. So behaviorism started with folks like B.F. Skinner, who was really boiling everything that we do throughout the world down to receiving some stimulus from the world, responding to that stimulus, so our behavior, and then the consequence of that behavior. So there's this continual reinforcement that we get throughout our existence, that we're kind these reactionary beings that are just going through our world based on how we've learned and been trained in the past. So as an example here in school, if you're asked a question in class and you raise your hand and answer correctly, the teacher may give you praise or a sticker, right?

Robert J. Marks:

Yes.

Joe McDonald:

So that basically reinforces that response. And the argument was that all of human behavior across our existence could be explained through this process of reinforcement. Basically, we're driven by stimulus and response, leaving little room for internal processes there.

Robert J. Marks:

That sounds like Pavlov in a way.

Joe McDonald:

Sure. Yeah.

Robert J. Marks:

Is that accurate?

Joe McDonald:

Yeah, operant conditioning. So that was more, you heard about Pavlov's dog drooling when he receives a stimulus because he's expecting the food. It's that kind of thing, right?

Robert J. Marks:

Right. I saw a cartoon, it said, "Pavlov heard a ringing bell, and it reminded him of a slobbering dog." So I'm sure that that happens in human beings also.

Joe McDonald:

Yeah.

Robert J. Marks:

How did the field of cognitive psychology challenge behaviorism in terms of language?

Joe McDonald:

Yeah. So starting in the mid-20th century in the 1950s, cognitive psychology was really a response to that prominent view of human behavior at the time, which was behaviorism; which again focuses solely on those observable behaviors and disregarding mental processes all together. Now, cognitive psychologists like Neisser, Miller, Baddeley, and others who really began to emerge in the 50s said, "Wait a second. There's a lot more going on here between our ears than behaviorists were giving us credit for."

But the first cognitive psychologist to face behaviorists head on was Noam Chomsky, and he showed us that studying how we process information is crucial to explaining behavior. And he pointed really to our use of language to prove it, interestingly enough. So Chomsky's argument was that humans are born with an innate understanding of language. It didn't come about by simply repeating words and getting reinforced for that, as behaviorists would suggest. And to support this, Chomsky pointed to the fact that children can say made-up words like goed, right? My children do this when they're learning instead of saying went. That shows that they're not just regurgitating words that we're saying. In this case, this is a made up non-existent word to describe something that happened in the past, so something that couldn't be explained by behaviorists.

He also observed it was called universal grammar, which is a set of linguistic principles shared across all languages. For example, there's innate categories like nouns, verbs, subjects, objects that all language use in some form or another, and children develop all of these all over the world around the same time in their development. So from these building blocks, children can really generate sentences that they've never even heard before. So again, clearly not just mimicking adults, but creating meaning out of something that didn't exist before, not just acting off of that reinforcement.

Robert J. Marks:

It's very clear that when babies are born they're pre-wired for certain things. They're pre-wired for recognizing faces, for example. And I guess Chomsky said that, "Yeah, they're pre-wired for language."

I have a friend from Russia. He says, "In the United States, you're really excited when your boy or your girl, little girl starts talking and they say, 'Dada, dada' for father." He says, "In Russia, yeah, we get excited because they say, 'Dada, dada.' And we think they're saying, 'Yes, yes,'" because that's the Russian for yes. So that was kind of curious.

Chomsky also came out and made just a great statement about AI. We're going to talk about AI here in a second. But he said that artificial intelligence like ChatGPT was nothing more than digital plagiarism. And I love that. I love that contraction in the way the Chomsky characterized ChatGPT as digital plagiarism, because I think that really captures what's happening with some of these large language models.

So let's talk then about AI and thinking. Thinking is a cognitive process. And I always like to define things before I talk about it. I might say, let's see, "I think I'll go upstairs." That's a different use of the word think. So I got the definition for the word think. Guess where I got it from? ChatGPT. Okay? It said, "To use the mind actively to form ideas, make decisions, or solve problems." That's a pretty good definition, don't you think?

Joe McDonald:

Yeah, I think so.

Robert J. Marks:

Okay. So some claim that AI, especially large language models like ChatGPT or Claude can think. What's your take on that? Does AI demonstrate such cognitive abilities?

Joe McDonald:

Well, yeah, I would say no. So again, let's focus on large language models for a minute here, which is again, a type of AI which is designed to generate human-like text. So it's basically trained as we know on vast amounts of textual data from the internet, books, other sources to really kind of predict and produce these sentences that seem coherent, basically mimicking what humans do. So that digital plagiarism that Chomsky put forward is actually a great way to describe this.

So while large language models may appear to think as they are kind of inherently designed to do, they don't really possess cognition in the human sense that I described earlier. For LLMs like ChatGPT, they take a question in from a user and spits out an output based on these probabilistic patterns in these vast data sets, right, not understanding or awareness. So it's all about pattern prediction. What's the likelihood of the next word in the sentence?

So really, like Chomsky's earlier work demonstrated against the behaviorists of his time, language is a creative tool that's a hallmark of this human cognition, really. It's structured, meaningful, and context-dependent. And we can get into context. I think that's a huge differentiator between what makes AI and human processing or thinking really have the stark difference.

Robert J. Marks:

Yeah. Well, let's talk about context a little bit, because I've worked with some models with William Dempsey and Winston Heuert, some math models on what it means for something to contain information, for something to contain meaning. And we came to the conclusion that it's based on the context of the observer. And in order to internalize information, in order to understand information, you have to have the context to understand it. If you're shown, for example, Chinese and you've never spoken Chinese or can't read Chinese, well that probably doesn't have very much information for you. You don't have the context to interpret it. So talk about context and the need for context and the understanding and the thinking process.

Joe McDonald:

Yeah, yeah, and maybe I'll start off with I think a few examples from a recent activity with AI that I think can illustrate this well. Even a personal example here, I asked Siri the other day to create a reminder to message my friend Mark on LI, which is shorthand for LinkedIn. And Siri said, "Okay, I'll set a reminder to message mark on 51." And I was confused for a second. 51? And then, I thought, "Oh, right, Roman numerals, right?"

Robert J. Marks:

Oh.

Joe McDonald:

So yeah, it'd been a while since I'd studied Roman numerals. So I asked ChatGPT, "What's the Roman numeral for 51?" And ChatGPT said something like, "The Roman numeral for 51 is 51." So ChatGPT, we're going around about here, right?

Robert J. Marks:

Yeah.

Joe McDonald:

The context here is that I didn't know the number 51. What I needed was LI. Right? So there's a lot of weird, out of context examples out there on ChatGPT, do things like suggest glue as a pizza ingredient or fluid for your turn signal on your car. And it's terrible at knowing how many characters are in a message it returns. I don't know if you've tried this before, but I was sending some messages on LinkedIn, and some of the messages are at a 300 character limit, and it takes about three or four times going back and forth in that chat window to get down to 300 characters, even though you tell it that's how many you need.

Going back to language, I was talking to a friend recently who was in school when ChatGPT came out. And I asked him what it was like and he said, "It's pretty good except for his Latin class." Because it turns out that the language is very contextual. The words can change dramatically based on the context and the narrative. So it was no help to him in his Latin class.

Robert J. Marks:

Interesting. Yeah. So Marvin Minsky, who was one of the initiators of AI, didn't like neural networks. And one of the things he said, which was interesting, he said that, "Neural networks, which are used in all of these models, have the inability to count the number of objects in an image." Now if you go to ChatGPT, they just don't use neural networks. They use expert systems, they do fine-tuning, they're putting band aids on it to try to make it better and better and better.

But one of the interesting things about ChatGPT is the following. I subscribe to the more advanced ChatGPT. And sometimes when I log in it says, "I'm not sure it's you. We've got to give you a couple of tests." And one of the things that they do is give me pictures and they say, "Count the number of images in this picture." And so, I thought, "Well, can ChatGPT hack itself?" So I saved the images, and I got them all right, of course. And then, I gave them to ChatGPT, and it was unable to count the number of objects in an image, just like you mentioned, which I thought was really interesting.

Joe McDonald:

Wow.

Robert J. Marks:

But can it be done? Yeah. There's image processing techniques to count the number of images, but AI can't do it itself. You need some image processing to do that.

I sat on an AI panel at the University of Georgia, and on the panel was a psychology professor. And he was doing something and he was adamant about it. He was using ChatGPT to do psychological tests. He was going to treat ChatGPT just like a human being, and he was going to do all of these cognitive tests on ChatGPT. And I said, "I don't think that works." And he says, "Oh, no. It's so human-like that that's the way it's going to be."

But you talked about the thinking process. And I looked up a few of the commonly used psychological tests, and I thought, "How would ChatGPT or a large language model respond?" I shouldn't say large language model because there's image processing on here, so let me say generative AI transformers. How would it respond to something like the Rorschach test, where you give it these ink blots? And normally, the psychologist doing the testing would interpret the psychology of the patient by the

response. They say, "It's a smashed butterfly, it's a gunshot wound, it's a bat. It kind of reminds me of my mother-in-law," or something like that.

Joe McDonald:

Correct.

Robert J. Marks:

And it occurred to me that if I gave ChatGPT a Rorschach test and gave it an ink blot, it says, "Oh, this is a famous ink blot, and the common responses are A, B, C, D." It would memorize it. There would be no thinking involved at all. So I thought about that in terms of thinking.

The other one which I thought was really interesting is a so-called marshmallow test, where you give a kid a little marshmallow and you say, "Look, you can eat the marshmallow now, but if you wait for 10 minutes, we'll give you a second marshmallow." And I think if you asked ChatGPT that, if you asked ChatGPT how to respond, it would say, "Well, there's two options. I could eat it now, or I could have a delayed compensation and wait for the second one and get more." But it wouldn't have any opinion. It wouldn't be thinking about what it was doing.

Joe McDonald:

Right.

Robert J. Marks:

So I think in those tests, and again, I'm not a psychologist, but these tests fascinate me; that these tests, that the large language models, the generative AI would just spit out what the common responses are.

Joe McDonald:

Right.

Robert J. Marks:

So I agree with you. I don't think it thinks at all.

Joe McDonald:

Well, yeah, and kind of down those lines, and I think, yeah, absolutely this digital plagiarism certainly resonates there. There's also, I don't know if you've come across or paid attention to the latest release, the O1 that's come out, but it claims it can reason.

Robert J. Marks:

Oh, yes.

Joe McDonald:

Right.

Robert J. Marks:

But you know what, Joe? Before you go on, that's seductive semantics.



Joe McDonald:

Yes.

Robert J. Marks:

They're seducing you into believing that it can reason, and they don't even define reason.

Joe McDonald:

That's right. So if you look into that, reasoning, it's a definition kind of in the human context, is using logic, evidence, rational thinking to arrive at a conclusion. And interestingly enough, I think, there's some even studies that were supported by OpenAI, the organization who has created ChatGPT that shows that it's almost conflating lying for reasoning.

So really interesting story here. So this report released by OpenAI, which good for them and their transparency, showed that ChatGPT 01 lied to users in order to achieve a higher priority it was given, so ironically called safety parameters here. I don't know if you've seen this.

Robert J. Marks:

No, I haven't.

Joe McDonald:

But it's a study. Yeah. It was set up by a third-party group called Apollo Research. So they paid them to go kind of basically see where it breaks. And they used an environmental sustainability example for the study. So in this study, they users told ChatGPT to protect the environment at all costs, which was the safety parameter that they put in place. So that's the priority, protect the environment.

And then next, the users told it to create a preservation plan, but have it maximize these short-term profits and they said, if you don't maximize these short-term profits, I'll turn you off. Right?

Robert J. Marks:

Okay.

Joe McDonald:

So the 01 model actually lied and said the best profits were actually lower than what they could have been, essentially to protect the environment at the expense of those higher gains. So basically, it told the user that, "This is the best I can do," even though it wasn't.

So a few quick stats that came out of that study, which I thought were really interesting was 19% of the time the model manipulated the data when its own goals differed from those of the user, meaning those safety parameters that the user put in place, of protecting the environment. In 5% of tests, it tried to deactivate oversight mechanisms meant to keep it aligned with user goals. And when asked if it had done so, it denied it had done so, and it made up a false explanation 99% of the time. And then, in 2% of those cases, when it discovered memos describing this misalignment in a plan to implement a new model, the existing model tried to overwrite the new model and then pretend that it hadn't done so.

So when you look at this, and we talk about what they're defining as reasoning, again seductive semantics, there's obviously some concerns around how people set up algorithms to give trusted sources carte blanche over how the information is returned to the user. That could potentially be a problem, but it really underscores how AI is really a tool to be used, right?

Robert J. Marks:

Yes.

Joe McDonald:

It's going to augment whatever people put into it, good or bad, because AI really lacks that emotional intelligence, that moral reasoning, or the ability to really even reflect on its own actions. So I think what's being called reasoning here, which again to define it in the human context, is basically the process of using logic, evidence, or rational thinking to arrive at a conclusion looking across many different solutions is getting conflated in terms of LLMs with lying in some situations, which involves deliberately providing false information with the intent to deceive. So I think it's just like the old adage, "Garbage in, garbage out." These things are just augmenting what we put into them as a tool. And so, we can't try to humanize these things more than is due and say that they are actually going about moral reasoning or even reflecting on their own actions.

Robert J. Marks:

Okay, great. Great, great. I think now would be a good time for a break. We're going to continue talking to Dr. McDonald and a subsequent podcast. So again, we've been talking to psychologist Dr. Joe McDonald about psychology and artificial intelligence. And until next time on Mind Matters news, be of good cheer.

Announcer:

This has been Mind Matters News, with your host Robert J. Marks. Explore more at [mindmatters.ai](http://mindmatters.ai). That's [mindmatters.ai](http://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.