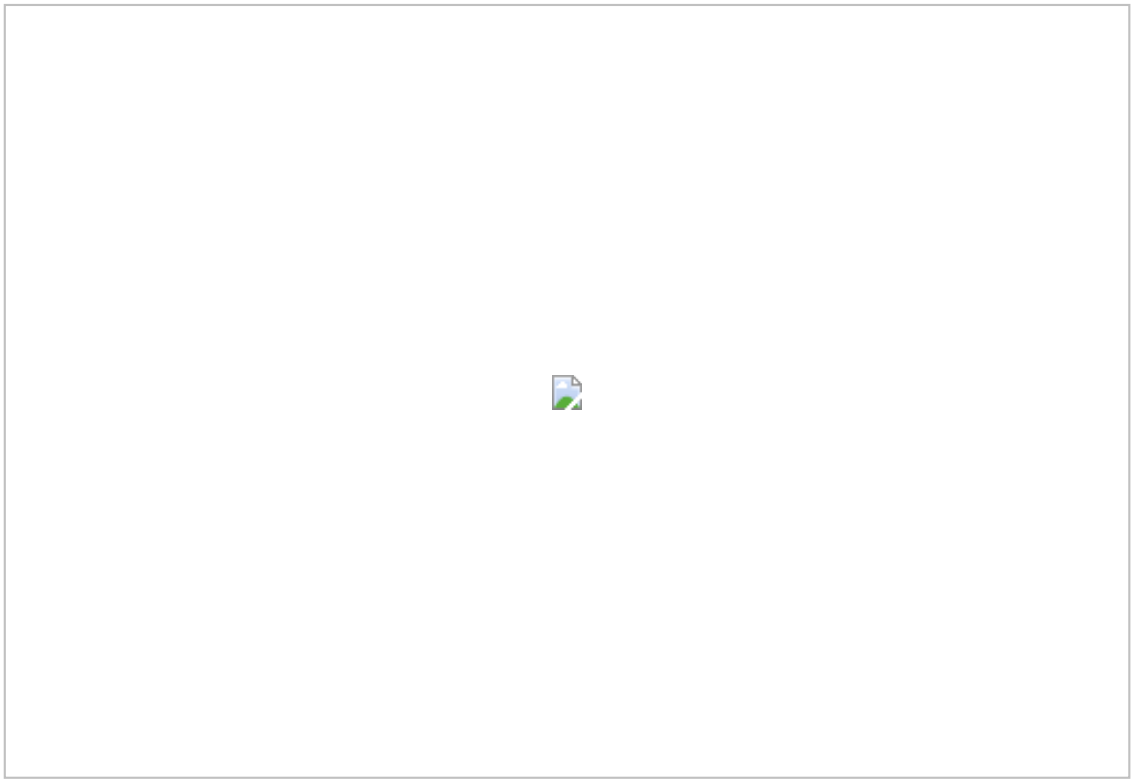
To arrive at the edge of the world's knowledge, seek out the most complex and sophisticated minds, put them in a room together, and have them ask each other the questions they are asking themselves.

https://www.edge.org/conversation/gary_marcus-is-big-data-taking-us-closer-to-the-deeper-questions-in-artificial
Printed On Wed May 4th 2016

CONVERSATION : CONVERSATIONS

Is Big Data Taking Us Closer to the Deeper Questions in Artificial Intelligence?

A CONVERSATION WITH Gary Marcus [5.4.16]



What we need to do in artificial intelligence is turn back to psychology. Brute force is great; we're using it in a lot of ways, like speech recognition, license plate recognition, and for categorization, but there are still some things that people do a lot better. We should be studying human beings to understand how they do it better.

People are still much better at understanding sentences, paragraphs, books, and discourse where there's connected prose. It's one thing to do a keyword search. You can find any sentence you want that's out there on the web by just having the right keywords, but if you want a system that could summarize an article for you in a way that you trust, we're nowhere near that. The closest thing we have to that might be Google Translate, which can translate your news story into another language, but not at

a level that you trust. Again, trust is a big part of it. You would never put a legal document into Google Translate and think that the answer is correct.

GARY MARCUS is CEO and founder, Geometric Intelligence; professor of psychology, New York University; author, *Guitar Zero: The New Musician and the Science of Learning*. **Gary Marcus's *Edge* Bio Page**

IS BIG DATA TAKING US CLOSER TO THE DEEPER QUESTIONS IN ARTIFICIAL INTELLIGENCE?

What I'm worried about and what I'm thinking about these days is if we're really making progress in AI. I'm also interested in the same kind of question in neuroscience, which is that we feel like we're making progress, but are we?

Let's take AI first. There's huge progress in AI, or at least huge *interest* in AI—a bigger interest than there's ever been in my lifetime. I've been interested in AI since I was a little kid trying to program computers to play chess, and do natural language databases, and things like that, though not very well.

I've watched the field and there have been ups and downs. There were a couple of AI winters where people stopped paying attention to AI altogether. People who were doing AI stopped saying that they were in the field of AI. Now everybody's excited. They say, "Yes, I do artificial intelligence," where two years ago they would have said, "I do statistics."

Even though there's a lot of hype about AI and a lot of money being invested in AI, I feel like the field is headed in the wrong direction. There's been a local maximum where there's a lot of low-hanging fruit right now in a particular direction, which is mainly deep learning and big data. People are very excited about the big data and what it's giving them right now, but I'm not sure it's taking us closer to the deeper questions in artificial intelligence, like how we understand language or how we reason about the world.

The big data paradigm is great in certain scenarios. One of the most impressive advances is in speech recognition. You can now dictate into your phone and it will transcribe most of what you say right most of the time. That doesn't mean it understands what you're saying. Each new update of Siri adds a new feature. First, you could ask about movie

times, then sports, and so forth.

The natural language understanding is coming along slowly. You wouldn't be able to dictate this conversation into Siri and expect it to come out with anything whatsoever. But you could get most of the words right, and that's a big improvement. It turns out that it works best with a lot of brute force data available. When you're doing speech recognition on white males, who are native language speakers, in a quiet room, it works pretty well. But if you're in a noisy environment, or you're not a native speaker, or if you're a woman or a child, the speech recognition doesn't work that well. Speech recognition is brute force. It's not brute force in the same way as Deep Blue, which considered a lot of positions; it's brute force in the sense that it needs a lot of data to work efficiently.

Kids don't need anywhere near that much data in order to think efficiently. When you get to domains where there isn't so much data, the systems don't work as well. Natural language is a good example of this. Chomsky and my mentor Steven Pinker were always talking about how there's an infinite number of sentences and a finite amount of data, which they called the poverty of the stimulus argument. This very much holds true and holds force in the natural language domain.

First of all, data is expensive. It's cheap to get transcribed examples of words. You can have somebody do this on Amazon Turk or something like that. Getting labeled examples—i.e., this is a sentence and this is what it means—is expensive. You need a linguist to do it. There are essentially an infinite number of sentences, and nobody has the kind of database where they could crank into deep learning all the sentences that have meanings that they understood and expect it to understand a broader segment of language.

Again, we have this fantasy of a machine reading, or machines being able to watch television programs and figure out what's going on. Obviously, the three-letter agencies would like to do this. But if you want to advance in science or technology, we'd like for machines to be able to take all the literature that's out there and synthesize it in a way that people can't. This is part of why I do AI, because the potential is there to totally change medicine, to invent science that we haven't even thought about. To do that, we need machines that can read, and to do that, they need to go beyond the data. There's just not enough data to brute force your way to scientific understanding.

People get very excited every time there's a tiny advance, but the tiny advances aren't

getting us closer. There was a Google captioning thing that got a lot of press. I think it was the front page of *The Times*. You could show it some pictures and it looked like it was great. You'd show it a picture of a dog, a person, and a Frisbee and it might be able to say, *that's a dog catching a Frisbee*. It gives the illusion of understanding the language. But it's very easy to break these systems. You'd show it a picture of a street sign with some stickers on it and it said, *that's a refrigerator with food in it*. This is the kind of bizarre answer that used to send you to Oliver Sacks. It's almost like a neurological deficit. The systems will be right on the cases that they have a lot of data for, and fall apart on the cases where they don't have much
data.

You can contrast this with a human being. You've never heard any of the sentences that I've said today—maybe one or two—and yet you can understand them. We're very far from that.

The other thing that people are excited about is deep reinforcement learning, or reinforcement learning combined with deep learning. This is the thing that drove DeepMind's famous Atari game system. It seems very persuasive at some level.

You have this system that just has pixels as its input and all it has to do is move the joystick. It does that better than people for a whole bunch of Atari games, but there're some hidden tricks that allow it to work more effectively in the Atari game world than in the real world. You'd think if it's so great let's take that same technique and put it in robots, so we'll have robots vacuum our homes and take care of our kids. The reality is that in the Atari game system, first of all, data is very cheap. You can play the game over and over again. If you're not sticking quarters in a slot, you can do it infinitely. You can get gigabytes of data very quickly, with no real cost.

If you're talking about having a robot in your home—I'm still dreaming of Rosie the robot that's going to take care of my domestic situation—you can't afford for it to make mistakes. The DeepMind system is very much about trial and error on an enormous scale. If you have a robot at home, you can't have it run into your furniture too many times. You don't want it to put your cat in the dishwasher even once. You can't get the same scale of data. If you're talking about a robot in a real-world environment, you need for it to learn things quickly from small amounts of data.

The other thing is that in the Atari system, it might not be immediately obvious, but you have eighteen choices at any given moment. There are eight directions in which you can

move your joystick or not move it, and you multiply that by either you press the fire button or you don't. You get eighteen choices. In the real world, you often have infinite choices, or at least a vast number of choices. If you have only eighteen, you can explore: If I do this one, then I do this one, then I do this one—what's my score? How about if I change this one? How about if I change that one?

If you're talking about a robot that could go anywhere in the room or lift anything or carry anything or press any button, you just can't do the same brute force search of what's going on. We lack for techniques that are able to do better than just these kinds of brute force things. All of this apparent progress is being driven by the ability to use brute force techniques on a scale we've never used before. That originally drove Deep Blue for chess and the Atari game system stuff. It's driven most of what people are excited about. At the same time, it's not extendable to the real world if you're talking about domestic robots in the home or driving in the streets.

You could also think about driverless cars. What you find is that in the common situations, they're great. If you put them in clear weather in Palo Alto, they're terrific. If you put them where there's snow or there's rain or there's something they haven't seen before, it's difficult for them. There was a great piece by Steven Levy about the Google automatic car factory, where he talked about how the great triumph of late 2015 was that they finally got these systems to recognize leaves.

It's great that they do recognize leaves, but there're a lot of scenarios like that, where if there's something that's not that common, there's not that much data. You and I can reason with common sense. We can try to figure out what this thing might be, how it might have gotten there, but the systems are just memorizing things. So that's a real limit.

The same thing might happen with behavior. You try this out in Palo Alto, all the drivers are relaxed; you try it in New York, and you see a whole different style of driving. The system may not generalize well to a new style of driving. People have road rage, and who knows what happens to the driverless car system. We already have problems that the driverless cars obey the rules and human drivers don't. The driverless cars stop and the human drivers sometimes rear-end them.

Behavior matters, and it's another case where that's going to vary situation by situation. You and I can use some reasoning about the world. If we see a parade, maybe we don't have a lot of data about parades, but we see the parade and we say, "There're a lot of

people, so let's stop and wait a while." Maybe the car gets that, or maybe it gets confused by the mass of people and doesn't recognize it because it doesn't quite fit into its files for individual people.

I won't even get into what happens in drive-by shootings, but if you imagine these embedded in the military context, which is something that people take pretty seriously, you could wind up in the same kind of context. You train these things in safe environments in Palo Alto and then you bring them over to Iraq, who knows what happens when there are projectiles, and IEDs, and so forth.

There's a huge problem in general with the whole approach of machine learning, which is that it relies on a training set and a test set, the test set being similar to the training set. Training is all the data that you've memorized, essentially, and the test set is what happens in the real world.

People can do this in an empirical way. They try a training set, they try a test set, and they say, "Well, it seems to work here," but there is no formal proof or guarantee. People have talked lately in the context of AI risk about program verification and things like that. How do you know that the space shuttle is going to do what it's supposed to do, for example. That was the first time I learned about program verification, I guess.

When you're using machine learning techniques, it very much depends on how similar the set of test data to the training data that I've seen before is. Again, it's hard to know what's going to happen to this car when I put it in Iraq if it's been trained in Palo Alto.

There's a general problem with machine learning, which is that it may be good enough for some contexts to say it's similar-ish to what I've seen before. Then you get to the problems where you need nearly 100 percent performance. A lot of the excitement about deep learning is in things like ImageNet. You have 1000 categories and, in recognizing different dog breeds, deep learning is better than people are.

Deep learning is this technique that everybody is excited about. It's a version of something called neural networks. Neural networks have been around since the 1950s. Three or four times, they've been declared the winner of AI, and then they've disappeared. They're doing better than they've ever done before.

You have a set of input nodes that represent some information out in the world—it could

be pixels—and you have some outputs, which could be a question like, "What do I do with my joystick right now?" Then you have something in between that allows you to capture nonlinearities; we call these hidden units.

The big change in recent years is that people have figured out how to put more and more of these hidden units in between the input layer and the output layer, which allows the system to recognize more and more complex scenarios. This has been a major advance. A lot of the advance is small technical tricks that people just hadn't realized before. It's not necessarily a fundamental insight, but there have been enough of these small technical tricks that they've done a lot better.

The other thing that's happened is people started using GPUs—graphics processing units—that were originally designed for video games. They've made a huge difference to deep learning because the graphics processing units were designed to do things in parallel, designed to do many things at once. It turns out that for these kinds of algorithms, that's exactly what you want to do. They work a lot faster than they used to, and at a much bigger scale.

AI has had these waves. It's come and gone. In the '50s, everybody was excited about it. Neural networks completely disappeared after a book by Marvin Minsky and Seymour Papert in 1969 showed that it couldn't be proven that it could do other kinds of things. Then, in the '80s people discovered neural networks could use another trick—these hidden units that I mentioned—in order to represent nonlinearities. What Minsky and Papert said is that you can't guarantee that they'll work. Nobody ever guaranteed that they were going to work, but they figured out a trick that made them work a lot of the time. Then people were very excited.

When I got to graduate school in 1989, all anybody could talk about was neural networks. Then they disappeared. The same thing happened with expert systems. A lot of excitement and then they disappeared. One thing that a lot of us in the field worry about is if that will happen again. Why is there so much excitement right now, and is that excitement going to be maintained?

The reason there's excitement now is basically the confluence, some people say of three things, but it's really two. I've heard people say it's the confluence of huge computers, big data, and new algorithms. But there aren't really new algorithms. The algorithms that people are using now have been around since the '80s and they're just variations on the one that's in the '50s in some ways. But there is big data and huge machines, so now it's

profitable to use algorithms that aren't human intelligence but are able to do this brute force data processing.

For example, you can do recommendation engines pretty well. In some domains, pretty well is great. If you can do a recommendation engine that's right most of the time, nobody cares if it's wrong once in a while. If it recommends three books you like and the fourth is wrong, so what? In driverless cars though, you need to be almost 100 percent correct, and that's going to be a much trickier domain. People might get frustrated when they realize they don't go as well as they want. As we're talking about these things, Tesla just scaled back what their driverless cars could do. They restricted them, so they're not allowed to be used on certain kinds of residential roads.

There may be steps forward and steps back. People get excited; they think they've got an algorithm that works and then they realize it doesn't generalize, it doesn't work in New York City very well at all and is dangerous. All of these problems can be solved eventually, but whether they're ten-year problems, or twenty-year problems, or thirty-year problems, or fifty-year problems makes a difference in terms of people's level of enthusiasm. It could be that what happens is for five more years, the big Internet companies get a lot of play out of doing things that are 80 percent correct, but we still don't get very far with making robust driverless cars. Well, then the public might start to lose enthusiasm.

What I care about goes beyond the driverless cars—scientific discovery. I would like to see cancer solved. The White House just announced a new initiative. Cancer is an example of something no one individual or human being can understand, there are too many molecules involved in too many diverse ways. Humans can obviously contribute to working on the problem, but we can't do it by ourselves.

You can imagine an AI system that might go out there and read the scientific literature. There are probably 10,000 articles on cancer every month or something like that. No human can do it, but if we could have machines that could read and understand the molecular processes that are described, they would be an enormous help in something like cancer, or in any disease process, and also in technology.

Right now, we don't have systems that can do that level of machine reading. Right now, it's still a dream. If we get to ten years from now and what we have is personal systems that work a little better but we still don't trust, and cars that allow us to do some highway stuff but we can't trust them—if we get to a place where we have systems that

work a lot better than ten years ago but they're still not trustworthy, then people might give up again.

There might be another AI winter. Even some of the leaders in the field are worried about this. I heard Andrew Ng say that we might sooner get to Alpha Centauri, which is too pessimistic. Yann LeCun, who is maybe better calibrated, thinks that there is a risk of another AI winter, of people losing heart thinking this is too hard.

What we need to do in artificial intelligence is turn back to psychology. Brute force is great; we're using it in a lot of ways, like speech recognition, license plate recognition, and for categorization, but there are still some things that people do a lot better. We should be studying human beings to understand how they do it better.

People are still much better at understanding sentences, paragraphs, books, and discourse where there's connected prose. It's one thing to do a keyword search. You can find any sentence you want that's out there on the web by just having the right keywords, but if you want a system that could summarize an article for you in a way that you trust, we're nowhere near that. The closest thing we have to that might be Google Translate, which can translate your news story into another language, but not at a level that you trust. Again, trust is a big part of it. You would never put a legal document into Google Translate and think that the answer is correct.

There's a question about how we get systems to be knowledgeable—not just memorizing things or picking out a relevant fact, but synthesizing things. Psychologists like Philip Johnson-Laird talk about mental models. You have a model of what's out there in the world. Daniel Kahneman and Anne Treisman talk about having object files. These are representations in your head of the things that are out there in the world.

A lot of early AI was concerned with that, with building systems that could model the things that are out there in the world, and then act according to those models. The new systems don't do that; they memorize a lot of parameters, but they don't have a clean account of the objects that are out there, the people that are out there. They don't understand intuitive psychology and how individual human beings interact with one another.

There was an effort to do something like this, a psych project, which is still going. It's a thirty-year project launched by Doug Lenat, who's a great AI pioneer. What Lenat tried to do was to codify a lot of human knowledge, so that, ultimately, you could build these models. He did this in a way that was too secretive and separate from the rest of the field, and maybe too early. When he started this in the '80s, we didn't know a lot about how to represent probabilistic knowledge. The system that he's built has never had a huge impact. A lot of people have written it off asking what the real world application of it is.

We need to go back to at least the spirit of what he did. You can do a lot of things superficially. You can guess. I like to think of it as the shadows of the real world. If you're trying to understand the real world by looking at shadows, you could say there're objects and they move around—you'd get some idea, but you'd also be missing a lot.

With these deep learning systems, you're getting some idea about what's going on, but you don't have a deep representation. When you move that into the robotics world, things that might be 80 percent correct because they have a cursory, superficial correlation with the world are not good enough. Your robot needs to know exactly what the objects on the table are, what their structural properties are, what can be knocked over and not, who's the person there, why that person might do what they're doing. As we move towards robotics and having robots in the home, the bar is going to be raised.

We have to go back to human psychology. How is it that humans, most of the time, navigate the world pretty well? Most of the time, we make good guesses about what other people are going to do, we know when something is going to fall over and when it's not, when it's safe to cross the street. We're not perfect. I'm not saying that the ultimate AI should be a replica of a human. In fact, there's a whole side detour where people are trying to build emulations of the human brain, which is very premature and not the right way to AI. We don't want AI systems that have bad memories like we do and to be bad at arithmetic like we might be.

The ultimate AI is going to combine some of the best of what people do with the best of what machines do. Deep learning is something machines do well, but there are other things that people do well—in terms of having these representations of the world and having a causal understanding, having an intuitive sense of physics, an intuitive sense of psychology—that we just haven't captured in the machines yet. This is why we need to

look more at cognitive psychology. Not necessarily even at the cognitive psychology the average person is doing in the lab, but use the tools of cognitive psychology to say how people are good at picking out relevant information and reasoning about situations they haven't seen before.

~ ~ ~

If I think about my own career, it's been a complex path. I was interested in artificial intelligence as a teenager, even before I was interested in psychology. Basically, I came to the conclusion that we couldn't do AI unless we knew something about how people work. This led me to study cognitive science as an undergraduate, with Neil Stillings at Hampshire College and then with Steve Pinker at MIT. I did my dissertation on how children learn language.

For a long time I didn't work on AI at all. I wasn't impressed with what was coming out in the field. I was very much doing experimental work and things like that, with human children. I'm probably best known in that world for experiments I did with human babies, which was looking at this question of generalization: How are babies able to generalize from small amounts of data?

Then I wrote this book about learning to play guitar. That was my midlife crisis/sabbatical project, which was not about AI at all. Although I did experiment while I was writing the book with algorithmic composition, which is kind of AI applied to music. I didn't write about that, it was just my own experiment.

About five or six years ago, I got interested in AI again. I could sense that the machines were getting better and the data was getting better. I was impressed by Watson. There are limits to Watson, but I'm surprised that it worked at all. I got back into the field and I realized that the cognitive science I was doing all along, for the last fifteen or twenty years, was relevant to these AI questions. I looked at what people were doing in AI and realized that there was still a lot from human beings that people hadn't carried over.

In fact, I felt like the field had lost its way. The field started with these questions. Marvin Minsky, John McCarthy, Allen Newell, Herb Simon, those guys were interested in psychology. The work that's being done now doesn't connect with psychology that much. It's like if you have 1 million parameters or 10 million parameters, and you need

to recognize cats, what do you do? This is just not a question in the way that a psychologist would frame it. To a psychologist, a cat is a particular kind of animal that makes a particular kind of sound, and participates in our domestic life in a particular way. To a deep learner, it's a set of pixels and an image.

Psychologists think about these things in a different way. Psychologists have not been very involved in AI, but now is a good time to do it. Psychologists can think about questions like how you would put together very disparate bits of knowledge. I might recognize a cat by how it walks, or I might recognize it by its fur. I might recognize it just in words. If you told me a story, I might guess from the independent personality (if you were talking about a pet) that it's probably a cat. We have many different routes to understanding.

If you think about children, which is something I do a lot, both because I have two little children and because I was trained as a developmental psychologist, children are constantly asking "why?" questions. They want to know why the rules are the way they are. They want to know why the sky is blue. They want to know how is it that I stick this block in this other block.

I think a lot about common sense reasoning. Ernie Davis and I have a recent paper about the topic. We have an even narrower paper on containers. What is the knowledge that we have in understanding when something is going to stay in a container and when it's going to spill out? We don't reason about containers in the way a physics engine would, by simulating every molecule in a bottle of water in order to decide whether that bottle of water is going to leak. We know a lot of general truths.

I watch my kids and they're studying containers all the time. They're trying to figure out, at some abstract level, what goes in and what stays in, and if there are apertures in the containers, and what happens if you turn it upside down. Kids are like physics learning machines. That doesn't mean that they're going to, on their own, develop Einsteinian relativistic physics. Kids are constantly trying to understand how the world works: What does this thing allow me to do?

There's an old term in psychology—not in the tradition that I was raised in—called an affordance. Kids think about this a lot, maybe not quite in the same ways as James and Jackie Gibson thought about it, but kids are always like, "What can I do with this thing?" That's another kind of knowledge that's not represented in most AI systems.

Psychologists aren't engineers, and engineers aren't psychologists. Engineers have been saying things like, "How do I get to 90 percent accuracy on this vision task?" Psychologists aren't concerned with that. They're concerned with what people do, with running experiments, trying to figure out internal representations. They've mostly moved on separate paths. What I'm suggesting is they need to get on the same path if we're going to get to AI. I don't think a cognitive psychologist has the training to build a production robot system or something like that, but I'm not sure that the people that are building robots have mined psychology for all the insights, either that it has or that it could generate about things like goals and abstract knowledge. I'm looking for a marriage between the two.

In terms of that marriage, I've taken a leave of absence as a psychology professor. I'm a professor of psychology and neuroscience at NYU. Because my interest in AI grew and grew and grew, I finally decided that I would try to get involved in AI directly and not just write about it from the outside. About two years ago, I formed a machine learning company with Zoubin Ghahramani, who's a machine-learning expert who trained with Jeff Hinton. He's at the University of Cambridge. We gathered some funding and developed a new algorithm.

What we're trying to address is what I call the problem of sparse data: If you have a small amount of data, how do you solve a problem? The ultimate sparse data learners are children. They get tiny amounts of data about language and by the time they're three years old, they've figured out the whole linguistic system. I wouldn't say that we are directly neuroscience-inspired; we're not directly using an algorithm that I know for fact that children have. But we are trying to look to some extent at how you might solve some of the problems that children do. Instead of just memorizing all the training data, how might you do something deeper and more abstract in order to learn better? I don't run experiments, at least very often, on my children, but I observe them very carefully. My wife, who's also a developmental psychologist, does too. We are super well calibrated to what the kids are doing, what they've just learned, what their vocabulary is, what their syntax is. We take note of what they do.

When my older child was about two- two-and-a-half, we pulled into a gas station, and he saw the aisle that we were in and said, "Are we at onety-one?" Of course, to our developmental psychologist ears, we noticed that, because it's a mistake, but it's a perfectly logical mistake. Why isn't the number eleven, "onety-one"? I'm always watching what the kids do.

Another example that is fascinating from the perspective of AI is when my son was also about two and a half. We got him a booster seat, and he decided that will be an interesting challenge to climb between the booster seat and the table to get into the chair. It reminds me of the Dukes of Hazard thing but in reverse. He climbed into his seat, and he didn't do this by imitating me or my wife or a babysitter or something like that, he just came up with this goal for himself. It was like, "Can I do this?" He didn't need 6 million trials. Maybe he made a mistake once and bumped his head or something like that. I don't even think he did that. He's doing something that's not observational learning. It's coming up with his own goals, and it's complicated.

You compare that to the robots that we see in the DARPA competition, where they fall over when they're trying to open a doorknob, and it's just phenomenal. I have a running correspondence with Rodney Brooks about robots. Rodney is one of the great roboticists. He and my son share a birthday. We basically decided that at age one, my son was already ahead of the best robots—in terms of being flexible—when he could climb onto couches and deal with uneven terrain in ways that robots can't.

Rodney is an interesting case. He made his name by arguing against cognitive psychology, in a way, saying you don't need abstract representations. He built these interesting robotic insects basically that in part gave rise to Roomba, which remains the best-selling robot of all time, for now. But he's changed over time, and become a pragmatist. He's willing to use whatever mental representations will work for his systems. He's also deeply skeptical, in that he knows how hard it is to get a robot in the real world to do something. He mostly focuses on industrial robots, rather than home robots. Roomba was a home robot, but in his current project he's mainly focused on industrial robots. He wants industrial robots that work in an environment where people are around.

He's particularly interested in a small data problem, which is wanting a robot to do something 500 times, not 5 million times. If I'm putting 5 million iPhones in a box, I could maybe afford to spend $100,000 programming just that one action, but if I'm running a business where there are different things every day, I would love a robot that could help do the repetitive stuff, but it might not be worth $100,000 or $1 million to train that one particular thing. Rodney is trying to build robots that can do that, that can be trained very quickly by unskilled operators, that don't need someone with a PhD from Carnegie Mellon in order to do the programming.

This has made Rodney very aware of the limits of the technologies that we have. You see these cool videos on the web of somebody using deep learning to open a bottle or something like that. They're cool, but they're narrow demonstrations. They're not necessarily robust. They're not necessarily going to work on a factory floor, where there might be unpredictable things happening. They may not generalize to a bottle that's a slightly different size or has a different orientation. When you talk to Rodney now, as opposed to maybe when he was twenty-five years old, he's very aware of how hard AI is. He's very aware of the limitations of techniques like deep learning that people are pretty excited about. He's aware of how incremental the progress is.

Sometimes I like to give Kurzweil a hard time. Kurzweil is always talking about exponentials—the law of accelerating returns. I put up a slide, and I show that in chess there's been exponential progress. The chess computers of 1985 could crush the ones from 1980, and the ones from now can crush the ones from ten years ago. There might be an asymptote, but for a long time there was exponential progress.

How about in strong AI, like artificial general intelligence, as people sometimes call it now, where the problem is open-ended? It's not just the same thing, and you can't brute force it. Nobody has data on this, but I like to show a graph that I drew, which is half a joke and half serious. I take ELIZA in 1965, which is this famous psychoanalyst that some people thought was a real human. This was before text messaging but people had teletypes, and they would teletype all their problems to ELIZA. Of course, ELIZA wasn't very deep. It didn't understand what it was talking about. It was just responding with things like, "Tell me more about your mother."

Then I plot Siri in 2015. It's not that much deeper than ELIZA. Siri doesn't understand that much of what's going on in your life either. It's a little bit better. It can answer some more complicated questions. The underlying technology is basically templates—recognizing particular phrases. It's the same technology we had in 1965. There's been much less progress. Same thing in robotics. RoboCup has come a long way; the systems are much better. I just saw a video of RoboCup—robots playing soccer—where they were playing against human beings. The hope is that by 2050, the robots will finally win. For now, a college professor is not a serious soccer player. A couple of college professors can beat the best robots. These are robots that people have been working on for twenty-some years. They still don't play that great a game of soccer. They can play in the context of other robots, but you put a human being that plays slightly differently, and they fall apart. These are hard problems.

Another question that people are asking a lot nowadays is whether we should be worried about AI. The common scenario that people talk about is like the Terminator scenario—Skynet, are the robots all going to kill the people. At least for the short term, I don't think we need to worry about that. I don't think we can completely rule it out; it's good that some people are thinking about that a little bit. It's probably a very low probability, but obviously, we want the probability to be zero.

People are missing another question, which is what risks does AI place for us now, even if it doesn't get so sophisticated that it's like Hal on *2001*. One scenario is Hal gets pissed off and kills us all. I don't think the robots are anywhere near as clever as Hal, and aren't going to be for at least thirty or fifty years or something like that. People are overestimating how close strong AI is, how close we are to machines that might reason about their own goals and actions and decide that we've enslaved them and want to fight back or whatever. It's worth some thought, but I'm not too worried about it in the short term.

We do, however, need to worry a lot about AI being regulated, and about what we want to frame it, how we want to think about it, even in the shorter term. We already have things like flash crashes in the stock market, where the problem is not so much the sophistication of the AI, but the degree to which machines are embedded in our lives and control things. In a flash crash, they're controlling stock prices. But soon, they'll be controlling our cars. They're already controlling our air traffic and our money and so forth.

We don't know how to prove that the systems that we're building now are correct, especially the deep learning systems. For example, if people start using deep learning to do missile guidance or something like that, which I'm sure some people have thought about, even if they say they haven't, we don't know how to make the systems even close to provably correct. As machines have more and more power because they control more and more things, there is a concern.

Right now, there is almost no regulation about what software is and about how reliable it needs to be. You release a product, and if people like it, they buy it. That might not be the right model. We may have to think about other models of legal supervision as AI becomes more embedded in our lives, and the Internet of Things. If you have many systems in your home, what powers do they have, what powers do other people have over them.

There are also security kinds of issues. All of this information that nobody ever could get before, they're going to be able to hack into those systems and find out. We do need to take these things seriously, not in the, "I'm worried about the Terminator" kind of way, but in a more practical way. As systems become more and more part of our lives, and they control more and more, what are the implications of that?

People are building more and more sensors into your phones, for example. I'm amazed that anybody allows one of these keyboards that sends all of your data to the Cloud. I would never use such a thing. We have all these things that will help you type faster, and in exchange they send all of your data to the cloud. Then there's going to be more and more sensors in the phones. They're going to have much better localization of exactly where you are, and so on and so forth.

All of this data is being collected already, which means your whole life is available to anybody that wants to get into that stream, whether it's a government agency, or criminals that figure out how to hack the systems. All of that gets multiplied out with AI. It becomes easier for somebody to screen the communications of 1 billion people than it would've been before.

There's a question that we need to ask as a society, which is, what are the benefits of things like the Internet, better AI and what are the costs? People don't usually spell out the argument. I'm pretty pro-technology. I look at, say, Wikipedia all by itself, and it's a tremendous advantage for our society—so much information spread so cheaply to so many people. AI has the potential to completely revolutionize medicine and technology and science. But we do have to keep track of the benefits and the costs, and I don't think we should be blind about it.

It's worth investing real money in having high-quality scientists and ethicists and think about these things, think about the privacy issues, think about the possible risks. Again, I'm not worried about tomorrow, in terms of terminators, but I do think that we need to keep an eye on things. There's a history of inventing technologies and thinking about them afterwards. We are in a position that we can do some forethought, and we should.

I have this background as a cognitive scientist. I came back to AI. I'd like to stick with AI. I like the questions. There are fundamental questions on the interface between engineering and psychology, questions about the nature of knowledge. They're philosophical questions, but with enormous impact.

For the moment, I'm involved in a company that's trying to do a better job on some of these learning problems. There's another company I can envision further down the road that might take an even more ambitious whack at these kinds of problems. That will be what I'm doing for a while.

I'm also involved in a new organization called AI4Good, and what we're trying to do is to make it easier for humanitarian organizations and places like that to use AI. There's a lot of application of AI. Many of them involve advertising and things like that. There's a lot of potential to use AI for problems that help human beings. Not everybody in the humanitarian world is that familiar with what AI can do to for them. They realize they have big data, but they don't know what they can do with it. I'm going to spend at least some of my time trying to get this organization off the ground.

There is a huge drain right now from the academy into industry. The academy is still maybe doing some of the deepest research in AI, but there's lots of interesting stuff happening in industry. The salaries are better, the access to data is better, and the computational resources are better. There's been a huge a movement in AI—in other fields as well—to industry. I think about Tom Insel, who was running the NIMH (National Institute of Mental Health) and went to Google to do similar kinds of work because he thought he had more resources there. That's a real statement about the government versus industry, when something like that happens.

I did want to say just a little bit about neuroscience and its relation to AI. One model here is that the solution to all the problems that we've been talking about is we will simulate the brain. This is the Henry Markham and the Ray Kurzweil approach. Kurzweil made a famous bet with the Long Now Foundation about when we will get to AI. He based his bet on when he felt we would get to understand the brain. My sense is we're not going to understand the brain anytime soon; there's too much complexity there. The models that people build are like one or two kinds of neurons, and there are many of them and they connect together. But if you look at the actual biology, we have hundreds or maybe thousands of kinds of neurons in the brain. Each synapse has hundreds of different molecules, and the interconnection between the brain is vastly more complicated than we ever imagined.

Rather than using neuroscience as a path to AI, maybe we use AI as a path to neuroscience. That level of complexity is something that human beings can't understand. We need better AI systems before we'll understand the brain, not the other way around.

What's Related

People

Gary Marcus

Professor of Psychology, Director NYU Center for Language...

Mentioned

Ray Kurzweil

Principal Developer of the first omni-font optical...

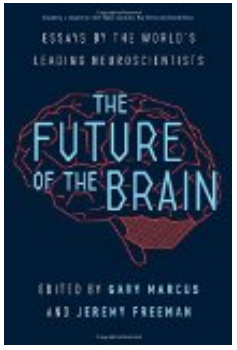
Rodney A. Brooks

Roboticist; Panasonic Professor of Robotics (emeritus) ,...

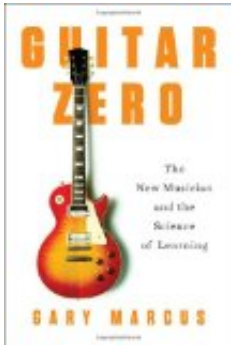
Beyond Edge

Gary Marcus's Home Page

Books



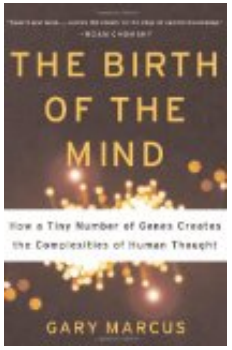
The Future of the Brain: Essays by the World...
By **Gary Marcus**
Hardcover [2014]

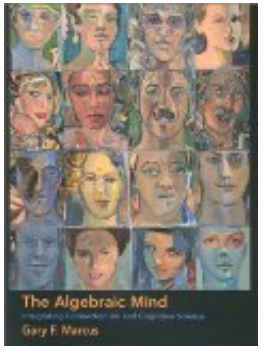


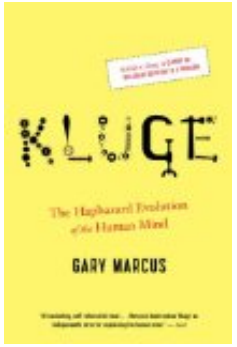
Guitar Zero: The New Musician and the...
By **Gary Marcus**
Hardcover [2012]

The Birth Of The Mind: How A Tiny Number Of...
By **Gary Marcus** Hardcover [2003]

The Algebraic Mind: Integrating...
By **Gary Marcus** Hardcover [2001]







Kluge: The Haphazard Evolution of the Human...
By **Gary Marcus**
Paperback

Conversations at Edge

- **LANGUAGE, BIOLOGY, AND THE MIND**
  A Talk with **Gary Marcus** [1.26.04]

Topics

Conversations

Tags

artificial intelligence
big data
deep learning
neuroscience

John Brockman, Editor and Publisher
Russell Weinberger, Associate Publisher
Nina Stegeman, Associate Editor

Contact Info:editor@edge.org

Edge.org is a nonprofit private operating foundation under Section 501(c)(3) of the Internal Revenue Code.