



“Although I do firmly  
believe that the  
brain is a machine,  
whether this  
machine is a computer  
is another question”

—RODNEY BROOKS



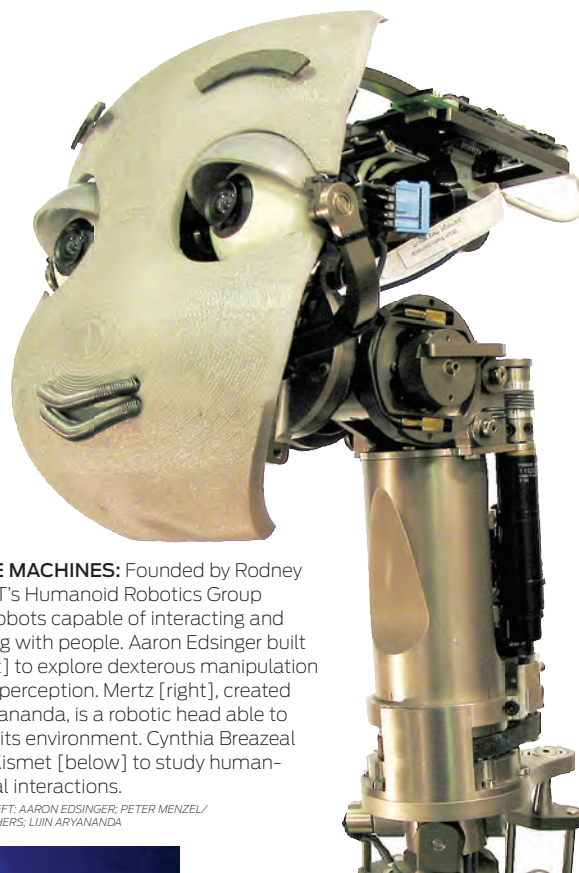
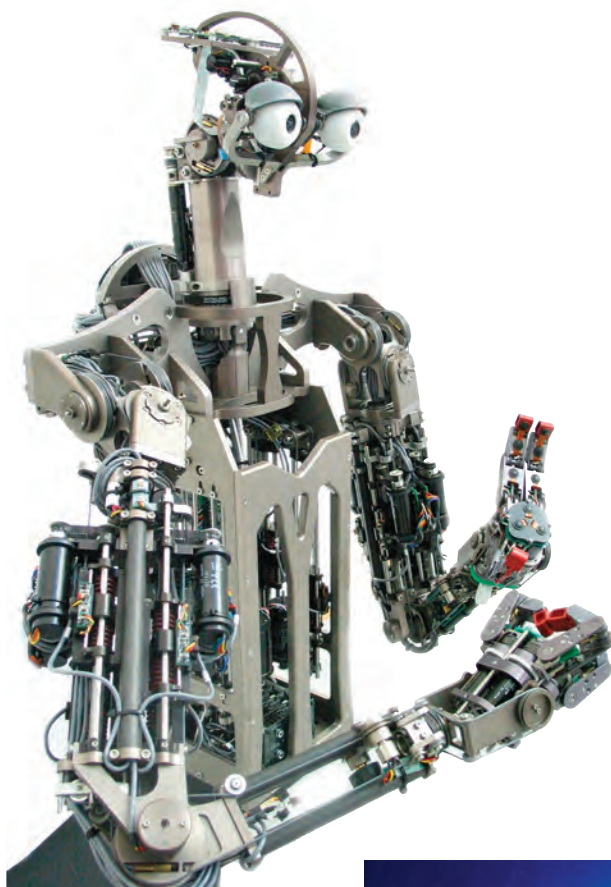


A POWERFUL ARTIFICIAL INTELLIGENCE WON'T SPRING FROM A SUDDEN TECHNOLOGICAL "BIG BANG"—IT'S ALREADY EVOLVING SYMBIOTICALLY WITH US  
BY RODNEY BROOKS

# I, Rodney Brooks, Am a Robot

AM A MACHINE. So are you. ¶ Of all the hypotheses I've held during my 30-year career, this one in particular has been central to my research in robotics and artificial intelligence. I, you, our family, friends, and dogs—we all are machines. We are really sophisticated machines made up of billions and billions of biomolecules that interact according to well-defined, though not completely known, rules deriving from physics and chemistry. The biomolecular interactions taking place inside our heads give rise to our intellect, our feelings, our sense of self. ¶ Accepting this hypothesis opens up a remarkable possibility. If we really are machines and if—this is a big *if*—we learn the rules governing our brains, then in principle there's no reason why we shouldn't be able to replicate those rules in, say, silicon and steel. I believe our creation would exhibit genuine human-level intelligence, emotions, and even consciousness.



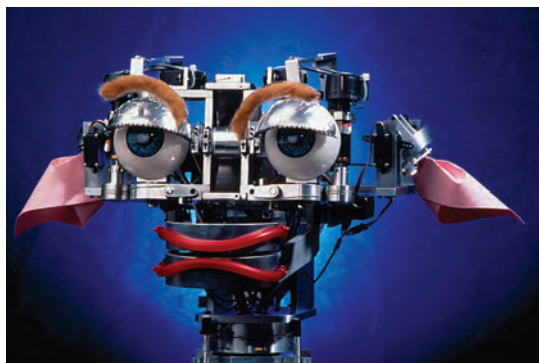


**SOCIABLE MACHINES:** Founded by Rodney Brooks, MIT's Humanoid Robotics Group develops robots capable of interacting and cooperating with people. Aaron Edsinger built Domo [left] to explore dexterous manipulation and visual perception. Mertz [right], created by Lijin Aryananda, is a robotic head able to learn from its environment. Cynthia Breazeal designed Kismet [below] to study human-robot social interactions.

PHOTOS, FROM LEFT: AARON EDSINGER; PETER MENZEL/  
PHOTO RESEARCHERS; LIJIN ARYANANDA

I'm far from alone in my conviction that one day we will create a human-level artificial intelligence, often called an artificial general intelligence, or AGI. But how and when we will get there, and what will happen after we do, are now the subjects of fierce debate in my circles. Some researchers believe that AGIs will undergo a positive-feedback self-enhancement until their comprehension of the universe far surpasses our own. Our world, those individuals say, will change in unfathomable ways after such superhuman intelligence comes into existence, an event they refer to as the singularity.

Perhaps the best known of the people proselytizing for this singularity—let's call them singularitarians—are acolytes of Raymond Kurzweil, author of *The Singularity Is Near: When Humans Transcend Biology* (Viking, 2005) and board member of the Singularity Institute for Artificial Intelligence, in Palo Alto, Calif. Kurzweil and his colleagues believe that this super AGI will be created either through ever-faster advances in artificial intelligence or by more biological means—"direct brain-computer interfaces, biologi-



cal augmentation of the brain, genetic engineering, [and] ultrahigh-resolution scans of the brain followed by computer emulation" are some of their ideas. They don't believe this is centuries away; they think it will happen sometime in the next two or three decades.

What will the world look like then? Some singularitarians believe our world will become a kind of techno-utopia, with humans downloading their consciousness into machines to live a disembodied, after-death life. Others, however, anticipate a kind of technodamnation in which intelligent machines will be in conflict with humans, maybe waging war against us. The proponents of the singularity are technologically astute and as a rule do not appeal to technologies that would violate the laws of physics. They well understand

the rates of progress in various technologies and how and why those rates of progress are changing. Their arguments are plausible, but plausibility is by no means certainty.

My own view is that things will unfold very differently. I do not claim that any specific assumption or extrapolation of theirs is faulty. Rather, I argue that an artificial intelligence could evolve in a much different way. In particular, I don't think there is going to be one single sudden technological "big bang" that springs an AGI into "life." Starting with the mildly intelligent systems we have today, machines will become gradually more intelligent, generation by generation. The singularity will be a period, not an event.

This period will encompass a time when we will invent, perfect, and deploy, in fits and starts, ever more capable systems, driven not by the imperative of the singularity itself but by the usual economic and sociological forces. Eventually, we will create truly artificial intelligences, with cognition and consciousness recognizably similar to our own. I have no idea how, exactly, this creation will come about. I also don't know when it will happen, although I strongly sus-

pect it won't happen before 2030, the year that some singularitarians predict.

But I expect the AGIs of the future—embodied, for example, as robots that will roam our homes and workplaces—to emerge gradually and symbiotically with our society. At the same time, we humans will transform ourselves. We will incorporate a wide range of advanced sensory devices and prosthetics to enhance our bodies. As our machines become more like us, we will become more like them.

And I'm an optimist. I believe we will all get along.

LIKE MANY AI researchers, I've always dreamed of building the ultimate intelligence. As a longtime fan of *Star Trek*, I have wanted to build Commander Data, a fully autonomous robot that we could work with as equals. Over the past 50 years, the field of artificial intelligence has made tremendous progress. Today you can find AI-based capabilities in things as varied as Internet search engines, voice-recognition software, adaptive fuel-injection modules, and stock-trading applications. But you can't engage in an interesting heart-to-power-source talk with any of them.

We have many very hard problems to solve before we can build anything that might qualify as an AGI. Many problems have become easier as computer power has reliably increased on its exponential and seemingly inexorable merry way. But we also need fundamental breakthroughs, which don't follow a schedule.

To appreciate the challenges ahead of us, first consider four basic capabilities that any true AGI would have to possess. I believe such capabilities are fundamental to our future work toward an AGI because they might have been the foundation for the emergence, through an evolutionary process, of higher levels of intelligence in human beings. I'll describe them in terms of what children can do.

■ The object-recognition capabilities of a 2-year-old child. A 2-year-old can observe a variety of objects of some type—different kinds of shoes, say—and successfully categorize them as shoes, even if he or she has never seen soccer cleats or suede oxfords. Today's best computer vision systems still make mistakes—both false positives and false negatives—that no child makes.

■ The language capabilities of a 4-year-old child. By age 4, children can

engage in a dialogue using complete clauses and can handle irregularities, idiomatic expressions, a vast array of accents, noisy environments, incomplete utterances, and interjections, and they can even correct nonnative speakers, inferring what was really meant in an ungrammatical utterance and reformatting it. Most of these capabilities are still hard or impossible for computers.

■ The manual dexterity of a 6-year-old child. At 6 years old, children can grasp objects they have not seen before; manipulate flexible objects in tasks like tying shoelaces; pick up flat, thin objects like playing cards or pieces of paper from a tabletop; and manipulate unknown objects in their pockets or in a bag into which they can't see. Today's robots can at most do any one of these things for some very particular object.

■ The social understanding of an 8-year-old child. By the age of 8, a child can understand the difference between what he or she knows about a situation and what another person could have observed and therefore could know. The child has what is called a "theory of the mind" of the other person. For example, suppose a child sees her mother placing a chocolate bar inside a drawer. The mother walks away, and the child's brother comes and takes the chocolate. The child knows that in her mother's mind the chocolate is still in the drawer. This ability requires a level of perception across many domains that no AI system has at the moment.

But even if we solve these four problems using computers, I can't help wondering: What if there are some essential aspects of intelligence that we still do not understand and that do not lend themselves to computation? To a large extent we have all become computational bigots, believers that any problem can be solved with enough computing power. Although I do firmly believe that the brain is a machine, whether this machine is a computer is another question.

I recall that in centuries past the brain was considered a hydrodynamic machine. René Descartes could not believe that flowing liquids could produce thought, so he came up with a mind-body dualism, insisting that mental phenomena were nonphysical. When I was a child, the prevailing view was that the brain was a kind of telephone-switching network. When I was a teenager, it became an electronic computer, and later, a massively parallel digital computer. A few

years ago someone asked me at a talk I was giving, "Isn't the brain just like the World Wide Web?"

We use these metaphors as the basis for our philosophical thinking and even let them pervade our understanding of what the brain truly does. None of our past metaphors for the brain has stood the test of time, and there is no reason to expect that the equivalence of current digital computing and the brain will survive. What we might need is a new conceptual framework: new ways of sorting out and piecing together the bits of knowledge we have about the brain.

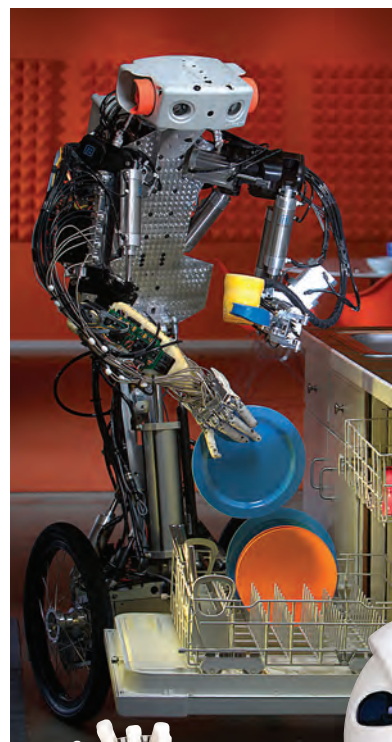
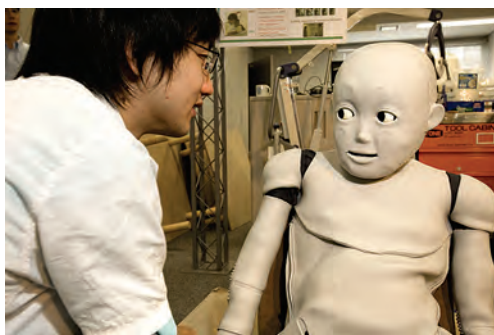
Creating a machine capable of effectively performing the four capabilities above may take 10 years, or it may take 100. I really don't know. In 1966, some AI pioneers at MIT thought it would take three months—basically an undergraduate student working during the summer—to completely solve the problem of object recognition. The student failed. So did I in my Ph.D. project 15 years later. Maybe the field of AI will need several Einsteins to bring us closer to ultraintelligent machines. If you are one, get to work on your doctorate now.

GREW UP in a town in South Australia without much technology. In the late 1960s, as a teenager, I saw *2001: A Space Odyssey*, and it was a revelation. Like millions of others, I was enthralled by the soft-spoken computer villain HAL 9000 and wondered if we could one day get to that level of artificial intelligence. Today I believe the answer is yes. Nevertheless, in hindsight, I believe that HAL was missing a fundamental component: a body.

My early work on robotic insects showed me the importance of coupling AI systems to bodies. I spent a lot of time observing how those creatures crawled their way through complex obstacle courses, their gaits emerging from the interaction of their simple leg-control programs and the environment itself. After a decade building such insectoids, I decided to skip robotic lizards and cats and monkeys and jump straight to humanoids, to see what I could do there.

My students and I have learned a lot simply by putting people in front of a robot and asking them to talk to the machine. One of the most surprising things we've observed is that if a robot has a humanlike body, people will interact with it in a humanlike way. That's one of the reasons I came to believe that to





**ANDROIDS ARISING:** Clockwise from bottom left: Honda's Asimo can dance and climb stairs and has even conducted an orchestra; the Actroid female humanoid was developed by the Japanese firm Kokoro and Osaka University; Chinese roboticist Zou Ren Ti, of the Xi'an Chaoren Sculpture Research Institute, sits next to his android twin [right]; Anybots, in Mountain View, Calif., is developing tele-operated mechanical servants; Toyota's Partner robots include little droids that play the



trumpet and the violin; Osaka University researchers built the CB2 robot to mimic the appearance and behavior of a toddler.

PHOTOS, CLOCKWISE FROM BOTTOM LEFT: HONDA; YVES GELLIE/CORBIS; INDIANA UNIVERSITY; ANYBOTS; TOYOTA; ANDRONIKI CHRISTODOULOU/WPN

build an AGI—and its predecessors—we'll need to give them a physical constitution.

At this point, I can guess what you're wondering. What will AGIs look like and when will they be here? What will it be like to interact with them? Will they be sociable, fun to be around?

I believe robots will have myriad sizes and shapes. Many will continue to be simply boxes on wheels. But I don't see why, by the middle of this century, we shouldn't have humanoid robots with agile legs and dexterous arms and hands. You won't have to read a manual or enter commands in C++ to operate them. You will just speak to them, tell them what to do. They will wander around our homes, offices, and factories, performing certain tasks as if they were people. Our environments were designed and built for our bodies, so it will be natural to have these human-shaped robots around to

perform chores like taking out the garbage, cleaning the bathtub, and carrying groceries.

Will they have complex emotions, personalities, desires, and dreams? Some will, some won't. Emotions wouldn't be much of an asset for a bathtub-cleaning robot. But if the robot is reminding me to take my meds or helping me put the groceries away, I will want a little more personal interaction, with the sort of feedback that lets me know not just whether it's understanding me but how it's understanding me. So I believe the AGIs of the future will not only be able to act intelligently but also convey emotions, intentions, and free will.

So now the big question is: Will those emotions be real or just a very sophisticated simulation? Will they be the same kind of stuff as our own emotions? All I can give you is my hypothesis: the robot's

emotional behavior can be seen as the real thing. We are made of biomolecules; the robots will be made of something else. Ultimately, the emotions created in each medium will be indistinguishable. In fact, one of my dreams is to develop a robot that people feel bad about switching off, as if they were extinguishing a life. As I wrote in my book *Flesh and Machines* (Pantheon, 2002), "We had better be careful just what we build, because we might end up liking them, and then we will be morally responsible for their well-being. Sort of like children."

**M**ANY OF THE advocates of the singularity appear to the more sober observers of technology to have a messianic fervor about their predictions, an unshakable faith in the certainty of their predicted future.

To an outsider, a lot of their convictions seem to have many commonalities with religious beliefs. Many singularitarians believe people will conquer death by downloading their consciousnesses into machines before their bodies give out. They expect this option will become available, conveniently enough, within their own lifetimes.

But for the sake of argument, let's accept all the wildest hopes of the singularitarians and accept that we will somehow construct an AGI in the next three or four decades. My view is that we will not live in the techno-utopia future that is so fervently hoped for. There are many possible alternative futures that fit within the themes of the singularity but are very different in their outcomes.

One scenario often considered by singularitarians, and Hollywood, too, is that an AGI might emerge spontaneously on a large computer network. But perhaps such an AGI won't have quite the relationship with humans that the singularitarians expect. The AGI may not know about us, and we may not know about it.

In fact, maybe some kind of AGI already exists on the Google servers, probably the single biggest network of computers on our planet, and we aren't aware of it. So at the 2007 Singularity Summit, I asked Peter Norvig, Google's chief scientist, if the company had noticed any unexpected emergent properties in its network—not full-blown intelligence, but any unexpected emergent property. He replied that they had not seen anything like that. I suspect we are a long, long way from consciousness unexpectedly showing up in the Google network. (Unless it is already there and cleverly concealing its tracks!)

Here's another scenario: the AGI might know about us and we know about it, but it might not care about us at all. Think of chipmunks. You see them wandering around your garden as you look out the window at breakfast, but you certainly do not know them as individuals and probably do not give much thought to which ones survive the winter. To an AGI, we may be nothing more than chipmunks.

From there it's only a short step to the question I'm asked over and over again: Will machines become smarter than us and decide to take over?

I don't think so. To begin with, there will be no "us" for them to take over from. We, human beings, are already starting

to change ourselves from purely biological entities into mixtures of biology and technology. My prediction is that we are more likely to see a merger of ourselves and our robots before we see a stand-alone superhuman intelligence.

Our merger with machines is already happening. We replace hips and other parts of our bodies with titanium and steel parts. More than 50 000 people have tiny computers surgically implanted in their heads with direct neural connections to their cochleas to enable them to hear. In the testing stage, there are retina microchips to restore vision and motor implants to give quadriplegics the ability to control computers with thought. Robotic prosthetic legs, arms, and hands are becoming more sophisticated. I don't think I'll live long enough to get a wireless Internet brain implant, but my kids or their kids might.

And then there are other things still further out, such as drugs and genetic and neural therapies to enhance our senses and strength. While we become more robotic, our robots will become more biological, with parts made of artificial and yet organic materials. In the future, we might share some parts with our robots.

We need not fear our machines because we, as human-machines, will always be a step ahead of them, the machine-machines, because we will adopt the new technologies used to build those machines right into our own heads and bodies. We're going to build our robots incrementally, one after the other, and we're going to decide the things we like having in our robots—humility, empathy, and patience—and things we don't, like megalomania, unrestrained ambition, and arrogance. By being careful about what we instill in our machines, we simply won't create the specific conditions necessary for a runaway, self-perpetuating artificial-intelligence explosion that runs beyond our control and leaves us in the dust.

When we look back at what we are calling the singularity, we will see it not as a singular event but as an extended transformation. The singularity will be a period in which a collection of technologies will emerge, mature, and enter our environments and bodies. There will be a brave new world of augmented people, which will help us prepare for a brave new world of AGIs. We will still have our emotions, intelligence, and consciousness.

And the machines will have them too. □



## EXPERT VIEW: Esther Dyson

### WHO SHE IS:

Commentator and evangelist for emerging technologies, investor and board member for start-ups; currently focused on health care, genetics, private aviation, and commercial space. Ran PC Forum conference until 2007; currently hosts the annual Flight School conference.

### THOUGHTS

"The singularity I'm interested in will come from biology rather than machines. We won't be building things; we'll be growing and cultivating them, and then they will grow on their own."

