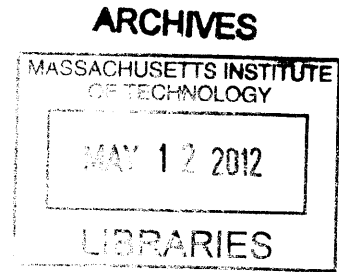


**Stronger:  
The Architects of a New Intelligence**

by

**Stephen Paul Craft**

Bachelor of Science, Computer Science  
Metropolitan State College of Denver, 2011



Submitted to the Program in Writing and Humanistic Studies  
in partial fulfillment of the requirements for the degree of

**MASTER OF SCIENCE IN SCIENCE WRITING**

at the

**MASSACHUSETTS INSTITUTE OF TECHNOLOGY**

September 2012

© 2012 Stephen Paul Craft. All rights reserved.

The author hereby grants to MIT permission to reproduce and to distribute  
publicly paper and electronic copies of this thesis document in whole  
or in part in any medium now known or hereafter created.

Handwritten signature of Stephen Paul Craft in black ink.

Signature of author: \_\_\_\_\_  
Graduate Program in Science Writing  
April 26, 2012

Certified by: \_\_\_\_\_  
Marcia Bartusiak  
Professor of the Practice  
Executive Director, Graduate Program in Science Writing

Accepted by: \_\_\_\_\_  
Thomas Levenson  
Professor of Science Writing  
Director, Program in Writing and Humanistic Studies

**Stronger:  
The Architects of a New Intelligence**

by

Stephen Paul Craft

Submitted to the Program in Writing and Humanistic Studies  
on April 26, 2012 in partial fulfillment of the requirements  
for the degree of Master of Science in Science Writing

**ABSTRACT**

The modern world is awash in technology, little of which amazes like artificial intelligence. Siri speaks to us out of our iPhones. Google answers our every question. Watson, IBM's *Jeopardy!*-playing supercomputer, is popularly perceived as a hair's breadth from Arthur C. Clarke's HAL 9000. But the truth of the matter is that all these technologies are far more *artificial* than they are *intelligent*. They are designed to give an impression of intelligence, a ruse at which many of them succeed. But none of them begins to approach the all-purpose intelligence of even a human child.

Siri and Watson and Google and all other existing AI is what researchers refer to as "narrow" — adept at one task only. But there is a small community of scientists who are working toward "strong" AI: a synthetic intelligence as flexible, as adaptable, and as genuinely intelligent as any human being. Strong AI is an innately interdisciplinary effort at the intersection of neuroscience, cognitive psychology, and computer science, among others. Each of these fields is coming to understand the challenge in its own way, and each has its champions.

Little is clear. Few agree on how best to build strong AI, and none can foresee its consequences. But one thing is certain: in constructing this new intelligence we will take our cues from an understanding of the human mind. And so the quest for strong AI ultimately becomes the quest to understand ourselves.

**Thesis Supervisor:** Marcia Bartusiak  
**Title:** Executive Director, Graduate Program in Science Writing

*For Stephanie,  
who believes in me*

It's a frosty January afternoon in Denver's art district, with sunlight filtering down through the clouds onto vacant sidewalks. On the south side of a nondescript building on Kalamath Street hangs a green corrugated metal door, ten feet tall and twelve wide; a faint whirring sound whispers behind it, a soft clicking. Then a pause, and the sound of a car ignition. Moments later the door rolls up and the car exits, turns left and heads north on Santa Fe Drive toward the skyscrapers downtown. As the door comes back down, a boxy white shape can be seen trundling through the shadows near the back of the garage. Welcome to the safest parking garage in Denver, guarded by Vigilus, the security-guard robot.

Vigilus looks more like an end table than an android. He's octagonal and a little over three feet tall. A speaker and a small rectangular computer display is mounted on his front panel between two columns of dots: sonar sensors, each about the size of a dime. Vigilus is undergoing one of his toughest tests yet this afternoon, patrolling the garage behind his creators' lab for eight consecutive hours with no human assistance. Over the course of the day he'll travel over three miles and make nearly sixty laps around the garage. He will constantly measure temperatures and track objects' locations, making special note of anything suspicious—open doors, objects in the thoroughfare, loitering humans.

Vigilus takes in information about his environment through the robot equivalent of eyes and ears: a video camera, his sonar system, and an array of

thermal sensors that line his front panels. But what allows him to differentiate between humans and cardboard boxes and automobiles, what makes it possible for him to make predictions about how best to navigate the garage, bears a striking resemblance to the thing that makes humans capable of performing the same tasks: Vigilus has a brain.

His brain is software, not cells, and it's simplistic next to the unfathomable intricacies of the human brain. But it does much of the same stuff, taking inputs from the world, interpreting them, storing memories, making predictions. Reasoning, in a sense. It allows him to be flexible and to adapt to his environment. But Vigilus's brain does something more profound than guide him around parking lots: it represents a return to an ideal.

More than sixty years ago, the founders of the field of artificial intelligence (AI) set out to design machines capable of thought—machines powered by the same processes that guide the human mind. As the decades wore on it became increasingly clear that such machines would be far more difficult to build than originally thought. The challenge eventually drove scientists away from their original goals and toward more pedestrian applications: Google, and Siri, and that friendly but infuriating robot voice on the airline reservation phone line.

Vigilus, while far from intelligent, demonstrates that modeling the human mind may still be the best idea we've had about how to make machines smarter, more flexible, more adaptable. A little more like their human creators.

\* \* \*

We find intelligent machines all over popular culture, and especially in science fiction—*2001: A Space Odyssey's* HAL 9000, the Terminator robot, *The Matrix's* malevolent agents. The frequency with which we encounter the idea of the intelligent machine, combined with an imperfect understanding of what it means to be intelligent, means that we assume too easily that the machines in the real world are, in some unclear way, just as smart as we are. We suppose IBM's chess-playing Deep Blue had a brain beneath all that plastic and silicon. We imagine that the iPhone's Siri is actually talking to us. In point of fact, however, we are almost as far from creating intelligent machines today as we've ever been.

It wasn't long after the invention of the computer before researchers started thinking about how to make them smarter. Concerted efforts began in the summer of 1956, when top scientists gathered at Dartmouth College in western New Hampshire with the ambitious intent of teaching computers how to learn like humans learn, and how to use human language—objectives they guessed wouldn't take more than three months to achieve. The conference was ultimately doomed by its lofty goals and naïve timetable. It never produced so much as a report.

The time wasn't wasted, though. The gathering launched artificial intelligence as a legitimate field of research: prior to 1956, those researching machine intelligence did their work in the absence of any distinct community. The very term "artificial intelligence" was coined at Dartmouth that summer. John McCarthy, a Stanford computer scientist, defined it as "the science and engineering of making intelligent machines, especially intelligent computer programs." It's as useful a definition as any, although its reliance on the ambiguously defined "intelligence" makes it a little knotty.

Viewed another way, AI is the science of making computers perform tasks that, were a human to perform them, would require some kind of intelligence. This encapsulates the pragmatic mindset of the overwhelming majority of AI researchers: those who seek to engineer machines that perform a specific task better, faster, or more efficiently than humans can.

McCarthy's original definition, though a little vague, is more ambitious; it identifies AI as a science explicitly concerned with replicating intelligence. Still, though, there's that troublesome word: *intelligence*. What does it mean to be intelligent? What are the qualities of an intelligent being? How do we know when something qualifies?

That's the question that Alan Turing set out to answer when, six years prior to the Dartmouth conference, he proposed an experiment intended not to define, but to

identify machine intelligence. Turing, a seminal figure in computer science, was convinced that computers could, in principle, be made to think, but was also concerned about the difficulty of defining intelligence. Rather than trying, he instead proposed his Turing Test, which more than six decades later remains one of the biggest unmet challenges in AI.

The test is simple. In it, two subjects, a human and a AI program, are sequestered behind closed doors. They interact with a second human, the judge, by exchanging messages—typed, so as to conceal the subjects’ identities. If the judge cannot distinguish the human from the computer, the AI is said to have passed the test.

The Turing Test has been a benchmark for intelligence for decades. Even today, scientists and programmers gather for the yearly Loebner Prize, awarded to the “most human-like computer,” judged relative to the other entrants. As a symbol of just how far we are from creating machine intelligence, the prize committee awards a bronze medal to the top performer. No one has ever won the silver, much less the gold.

The Loebner Prize uses the Turing Test as its criterion for success, but the Test illustrates a fundamental problem in how most AI researchers approach the problem of building intelligent machines: it requires the machine to deceive the judge. For example, an obvious question to ask might be, “What is 1,149,485 times 412,403?” The computer can calculate the answer (474,051,062,455) in microseconds, while the



human might require several minutes to work it out by hand. So in order to appear “human-like,” the computer must wait a few minutes before answering. A more straightforward approach might be to ask, “Are you a computer?” In this case the computer *has* to lie. It’s built into the rules of success.

More importantly, though, the Turing Test is inadequate because it allows for mere *simulation* of intelligence; if the AI gives sufficiently convincing answers, by Turing’s definition it can be said to be thinking even though we know it’s not, because we wrote the software and we know exactly what is—and what is not—going on behind the scenes.

Take Eliza for example. Eliza was the first “chatbot,” a computer program designed to interact with humans not in computer code or with punch cards but in old-fashioned human language—in precisely the format of the Turing Test. She was written in the mid-1960s by Joseph Weizenbaum, an MIT computer scientist, who designed her to speak like a psychotherapist. Here is a typical conversation with her:

<b>Eliza:</b>	Hello. How do you do. Please state your problem.
<b>Patient:</b>	My father never loved me.
<b>Eliza:</b>	Tell me more about your family.
<b>Patient:</b>	My parents just got divorced.
<b>Eliza:</b>	Your parents just got divorced?
<b>Patient:</b>	Three weeks ago.
<b>Eliza:</b>	Does that have anything to do with the fact that your father never loved you?
<b>Patient:</b>	It just got me thinking about our relationship.
<b>Eliza:</b>	Lets discuss further why your parents just got divorced.

She sure *seems* smart. But it doesn't take much for the illusion to break down.

**Eliza:** Hello. How do you do. Please state your problem.  
**Patient:** Doesn't that imply I only have one problem?  
**Eliza:** You say doesn't that imply you only have one problem?  
**Patient:** ...Yes? Isn't that what I said?  
**Eliza:** You seem to be quite positive.  
**Patient:** Um... thank you, I suppose.  
**Eliza:** We were discussing you—not me.  
**Patient:** This is a strange conversation.  
**Eliza:** I'm not sure I understand you fully.  
**Patient:** You and me both.

The patient makes an attempt at a joke (“Doesn't that imply I only have one problem?”) and the simulation immediately falls apart, because Eliza isn't programmed to respond to jokes.

Even at her most convincing, Eliza is simply a parrot, a mimic—there's no intelligence here. She doesn't need any knowledge about her interviewer, or the world, or about anything at all in order to pull off her charade. All she does is pick out the important words or phrases from the conversation (for example, she knows that “father” is a member of a “family”), wrap them up in a different sentence or a question, and throw them back at her interrogator. At her best she is convincing for only a short period of time; at her worst, she's comically obtuse.

Cleverbot, a more advanced modern chatbot, is subject to the same weaknesses. When researchers at Cornell pitted two copies of Cleverbot against each other, the conversation quickly degenerated into nonsensical theological ramblings,

one Cleverbot accusing the other of being a “meanie,” and the latter claiming to be a unicorn. Hilarious, sure, but not exactly smart.

By its very definition, the Turing Test doesn’t require any intelligence, only a convincing front. John Searle, a philosophy professor at the University of California at Berkeley, proposed a thought experiment in 1980 that he says illustrates the divide between the simulation of intelligence and actual understanding. In the experiment, which he calls the “Chinese Room Argument,” Searle sits in a room with a manual of the Chinese language, which provides a response to any possible question or statement in Chinese. Searle, who can’t tell Chinese from “meaningless squiggles,” receives notes in Chinese slipped under the door by native Chinese speakers. He looks up the characters in his manual, copies the appropriate response down on a slip of paper, and pushes it back under the door. The Chinese-speakers outside have no way of knowing that Searle, sitting in the room, doesn’t have the slightest clue about the substance of their conversation.

Of course, this is just another way of framing the Turing Test: Searle, in the thought experiment, is a computer, and the manual is the program he runs. His point is that simulated intelligence does not imply understanding. No program, he suggests, no matter how complicated or how comprehensive, can ever bestow a mind upon a computer.

Many hope to prove Searle wrong, but AI is such a new science that most researchers can't even agree on a way forward. AI has moved beyond chatbots, but even our most complex programs are still only superficially intelligent. Take Deep Blue for example. For decades, researchers thought that chess was the ultimate intelligence test. In 1997, after IBM's supercomputer defeated world-champion chess grandmaster Garry Kasparov in a six-game match, the New York Times hailed "the coming days of ascendant computers" and described Deep Blue's "personality" during the match as exhibiting "qualities of scrupulous care, unshakable calm, and remarkable powers of concentration and endurance."

In the following weeks and months, however, cooler heads prevailed. Deep Blue, for all its chess-playing prowess, was no more intelligent than any other computer—just a lot faster. More importantly, it didn't play chess like a human, with creativity and imagination alongside strategy and calculation. It did instead what computers do, evaluating 200 million moves per second and choosing the one most likely to get it a little closer to checkmate.

This is not to suggest what IBM's engineers accomplished was not difficult, but it was a feat of engineering, not of science: Instead of designing a computer that could think and therefore play chess, they simply designed a computer that could play chess. They concentrated on the ends, not the means, and in doing so solved a fundamentally easier problem.

The same tactic was visible in IBM's more recent effort at beating humans in games: Watson, the *Jeopardy!*-playing supercomputer. *Jeopardy!* seems like a game only a human could play, with awkwardly posed trivia questions, puns, and an incredibly wide range of material. Watson knew a lot of answers his human competitors did not, and doubtless appeared intelligent to many in the audience. But his designers know better, and a peek behind the scenes reveals more brute force: a vastly complex and nuanced search engine, but a search engine nonetheless.

"Whenever we understand how something works," says Patrick Winston, an AI researcher at MIT, "it loses its mystery. If I say something that sounds astonishingly smart to you, you'll think me smart until I explain how I had the thought, and then it will seem much diminished" — much as Eliza can seem intelligent until you understand how she picks and chooses her words.

The underlying problem with such systems is that they have no idea of how they fit in the world, no broader context. Watson and Deep Blue know *only* trivia or chess. Eliza's context is precisely the sum of what the user contributes to the conversation and what the programmer has pre-loaded into Eliza's memory. It's easy to build something that handles one topic well, but when you try to push that system outside its area of expertise, it breaks down.

This idea, that AI systems are constrained by what their programmers put into them, is at the heart of what researchers call "narrow AI." But there is an alternative:

“strong AI,” a term that John Searle coined despite the fact that he doesn’t believe it’s possible. Theoretically, strong AI would think like a human thinks, adapt to new problems, learn new skills, improve itself, and be flexible in ways traditional computers are completely incapable—as Searle puts it, a computer that has a mind “in exactly the same sense that human beings have minds.” For all intents and purposes strong AI *is a mind*, only built from silicon and computer code instead of neurons and synapses.

It’s hard not to hyperbolize when imagining the applications or abilities of such a computer. It would combine the efficiency and speed of traditional computers with the flexibility and adaptability of the human brain. In theory, given mechanisms for perceiving and interacting with the world, such a machine would be able to do anything a human could do, from tying a shoe to doing original scientific research.

The most ambitious and imaginative AI researchers believe that strong AI would be capable of solving some of the most difficult problems humans face, problems that arise from staggeringly complicated sets of interconnected variables unfathomable by the human brain. Such a machine could end poverty, or hunger, or disease. Such a machine could end war.

But back to earth; the reality isn’t nearly so impressive. Instead of grappling with such fundamental and intractable questions as *what is the mind*, most AI researchers take the comparatively simple approach of solving specific tasks. How do

we find things on the internet? How can we get a driver from Cincinnati to Des Moines? How can we win at *Jeopardy!*? And how can we make computers do these things faster and more efficiently than a human can? All existing artificial intelligence is narrow—*all* of it, even the most sophisticated and intelligent-seeming programs out there.

But there is a small and dedicated community of researchers who seek to advance the field in fundamentally new ways, scientists who are less interested in the narrow applications of AI than they are in understanding the mind, and strengthening AI based on that model. They concentrate on foundational questions and make substantive contributions to AI as a science while simultaneously expanding our understanding of what makes humans human.

Their approaches vary widely. Some think that biology is the way forward, that a neuron-by-neuron recreation of the brain is the surest way to produce intelligence. Others ask questions about what makes humans different from other intelligent species, and try to engineer systems that reproduce those differences. Some scientists think that describing the mind mathematically will provide a precise blueprint for intelligence. And yet others take cues from a variety of places, the pragmatism of business and the hard math of cognitive psychology and the rigid determinism of computer science.

But all of their research shares a common ancestor: the human mind. They believe it is the best model we have for flexibility, for adaptability, for generalized intelligence capable of surviving and growing in a complex world. We are the result of millions of years of evolutionary progress and the best (and only!) example we have of a system that gives rise to real intelligence, even if we can't say exactly what that means.

Their quest to build a smarter computer is, at its core, the quest to understand the human mind.

\* \* \*

At this very moment, photons reflected off the letters on this page are entering your eyeballs through your pupils, impacting your retinas, and sending electrochemical signals down your optic nerves and into your skull. Those signals start a chain reaction that jolts through your brain, racing from the visual cortex in the occipital lobe to a small region near the back of your head called Wernicke's Area, where written and spoken language are processed.

Your memory is activated. Your prefrontal cortex jumps to action to perform feats of logic and abstraction. Working in concert, your brain's regions recall the letters and words you read and form pictures of their meaning. Those pictures link together to form concepts, which link together to form ideas, which link together to form coherent thought. It is in part this sequence of linking, this spontaneous



networking both abstract and corporeal, that gives rise to intelligence. And it's all happening right now, this very moment, as you read these words.

The networks in the brain change over time, strengthened by repetition or weakened by disuse. The brain's ability to modify its own structure—to “self-organize”—is what neuroscientists call *plasticity*. Imagine how a school of fish looks in motion. No single fish decides where any other fish should be. Instead each fish obeys a few simple rules, instincts such as staying close, but not too close, to the fishes nearest it; swimming in the same general direction as those neighbors; and trying to match the other fishes' speed. With those rules in place the school takes on a life of its own—becomes beautiful, even. And it can behave intelligently, thwarting much larger predators through the concerted effort of myriad individuals.

The brain self-organizes in much the same way: a large number of small components arranged in a way that makes the whole smarter than any of its individual constituents. The fundamental units of the brain—the “fishes”—are neurons and synapses; neurons are cells that send and receive electrical impulses to and from each other, and synapses are the connections between neurons. No one neuron in your brain is in charge. But neurons can change their connections based on how other neurons in their neighborhood behave when information comes in through your eyes and ears and skin—much as the shape of the school of fish

changes in response to the presence of a predator. How the brain changes, how quickly, and how much, is plasticity, and it's how we learn.

While those networks in your brain are the best clue we have about where intelligence comes from, we simply don't know enough to recreate them. But Chi-Sang Poon, a biologist and an engineer with the Harvard-MIT Division of Health Sciences and Technology, has an idea about adapting what we know about building computers—which we're pretty good at—to model brains at the smallest scales. If we can do that, he thinks we may have a better shot at building fully functional, full-size, synthetic brains.

Poon is a biologist *and* an engineer—a point he makes emphatically. He seeks to bridge the divide between those sciences in very much the same way as he seeks to bridge the divide between the analog and the digital, between the brain and the computer. His lab on the second floor of MIT's Building E25 reflects this duality, with voltmeters and needle-nose pliers littering one workbench, cotton swabs and clean Petri dishes covering another.

His lab consists of four long, narrow rooms, each lined by workbenches. In the first room immediately adjacent to Poon's office, a collection of surgical instruments—scalpels, tweezers, forceps—is arrayed neatly on a white sheet, and a six-foot tower of electronic equipment stands in a corner. Wires trail out from this tower and attach to a collection of machinery on an adjacent table; Poon explains that

this is where he and his team do their biological experimentation. The pair of electrodes resting on the table spend much of their time plugged into rat brains.

In the next room over, wires hang down from an oscilloscope on a shelf, their leads connected to a circuit board on the workbench. This jumble of wires and circuitry is the centerpiece of Poon's research—a prototype of his “synaptic chip.” It resembles a three-layered hors d'oeuvre, with the chip itself resting atop two larger circuit boards like a dollop of caviar on a cracker. The synaptic chip—the caviar—is about an inch square. In its heart rests one square millimeter of electronics, a few hundred transistors that mimic the circuitry of the brain.

Designing this tiny facsimile has been no small task. The brain is an almost impossibly complex tangle, a network so complicated it makes a traditional computer's central processing unit (CPU) look like a child's toy. But Poon is starting small, examining the brain at the very smallest scales. He's interested in learning how we learn, which begins with just a single connection in the brain.

The brain contains an incredibly vast number of neurons and synapses: some hundred billion of the former, and about a hundred trillion of the latter—about as many synapses as there would be stars in 250 Milky Way galaxies. As we experience things, the neuron/synapse network gets strengthened in some places and weakened in others. Strong networks roughly equate to learned behavior or memory, so one pattern of connections in your brain might represent your understanding of English.

Another might be a recollection from your youth. The network related to language expands as you learn new words; the network representing the youthful memory weakens as you get older. In other words, it's plastic—it can change.

Recreating these spontaneous networks is the problem Poon has set out to solve. Until now, plasticity has been a characteristic strictly limited to brains: traditional CPUs (actually, all computer chips) are hard-wired and unchangeable, which is why CPUs don't learn. But Poon is using the best of both worlds: transistors, the building blocks of computers, arranged to emulate the shifting neural networks of the human brain. Poon is building computer chips that can learn.

His synaptic chip emulates the brain's plasticity, but uses fast, efficient computer components instead of neurons and synapses. Many other attempts at simulating brains at low levels have been done in software, using supercomputers to simulate the functions of neurons and neural networks. But even the fastest modern supercomputers can't mimic the brain's countless networks in real time.

Consider catching a baseball. The mathematics behind catching that ball, from tracking its flight with your eyes to moving your hand to intercept it—including all the brain-to-muscle signals to extend your arm, open your fingers, maintain balance, and so forth—are exceedingly complex, a huge set of simultaneous differential equations. A supercomputer has to formulate and solve all those equations hundreds of times a second to perform the same basic task—a slow and difficult process even

for a computer, whose native task is doing math. This traditional approach to AI uses “algorithms to explain the mind and the brain,” says Poon. “It’s an engineering or computer-science rendition of the brain—it’s highly simplified and highly reduced.”

The brain, on the other hand, doesn’t have to solve any equations, because those equations’ solutions are built into its neuronal networks. We’ve caught thousands of balls, and each time we repeat the action (or even think of the action) we strengthen the network associated with it—a network that literally embodies the math behind the physical act. Poon is building chips that learn in precisely the same way. It’s a technique that bypasses the rote calculation of traditional CPUs and may result in much faster computation.

For now, though, Poon’s chip doesn’t actually *do* a whole lot. It is a proof-of-concept—mathematical evidence that it is possible to mimic the behavior of the brain at the smallest scales. When we use our brains, the electricity that races through them sticks to very specific levels—a certain mathematical curve. If electricity is the currency of the brain, this curve is the dollar, the cent: the fundamental foundation on which everything else is built. These very specific electrical currents activate neurons, and lead inexorably to the spontaneous networks that eventually give rise to intelligence. Scientists have long known of this curve, but Poon is the first to reproduce it dynamically with his synaptic chip. For a guy working on building a brain from the bottom up, this is a crucial first step.

Still, Poon faces a long road. His prototype emulates just a single synapse, so he will need to add thousands more before he can replicate even the smallest piece of the human brain. But the biggest unknown is that scientists have very little idea how the brain's equation-solving networks are made, nor what they look like when fully formed. Poon theorizes that the way around this is to allow the chips to self-organize in the same way the brain does. Put the fundamental building blocks in place, he suggests, and let the chip work out the rest on its own. Whether intelligence will arise from such a chip is hard to say, and Poon himself concedes that such a result is a long way off. His challenge is not unlike the challenge faced by many other researchers working toward strong AI: we don't know where to start, so we're not sure where we'll end up.

Poon is interested in solving many of the same problems that more traditional AI researchers are working on, but he believes their approach is wrong. He thinks that any method that tries to emulate behavior, instead of the system that gives rise to that behavior, is doomed to failure. "We already know how neurons work," he says. "But that's not enough—we need to go a step further, to build a truly intelligent synapse that really mimics the intelligence behind [the brain]." With his synaptic chip, Poon has found a window into plasticity, one of the most basic elements of intelligence. With that in place, he says, "we have the basis for building an intelligent machine."

\* \* \*

For centuries, physicists have described the universe in mathematical terms. Newton used the expression  $F=MA$  to describe how a force ( $F$ ) acting on an object is equal to the mass ( $M$ ) of that object multiplied by its acceleration ( $A$ ), crystallizing the science of motion in three letters. Similarly, in 1905, Albert Einstein famously distilled the very nature of matter, creating what is arguably the most recognizable equation in the world:  $E=MC^2$ . Here each piece of matter in the cosmos holds an energy ( $E$ ) that is equivalent to the mass of that matter ( $M$ ) multiplied by the speed of light ( $C$ ) squared. Einstein's simplistic equation belies a profound elegance, and it revolutionized the way scientists conceive the universe.

Brain science, while younger than physics, is not exempt from this mathematical nature. Noah Goodman, a Stanford researcher who studied math and physics before turning to cognitive science, seeks to exploit this fact. He looks not at the biology of the brain but at its functions—understanding, memory, perception—and describes them mathematically. In doing so he neatly sidesteps our incomplete understanding of the physical structure of the brain.

Goodman's work doesn't leave a lot of room for mystery. The brain sure *seems* like a mysterious place: Strange things come burbling out of the subconscious, like songs popping into our heads unbidden, or an inexplicable craving for Twinkies at breakfast. Furthermore, it's not even possible to predict what our brain is going to do

with a constant input. You can look at a picture of a seascape every morning for a week, and while you might find it soothing on Monday, on Thursday the watery image might just make you rush to the bathroom.

Contrast this with computers, which are rigidly deterministic: given an input, you can always predict the output. Any deviance from the expected result happens not because the computer is feeling capricious or uncooperative, but because a programmer or engineer somewhere made a mistake. Given a sufficiently detailed description of a computer's software and its circuitry—a good manual—it is always possible to predict its behavior, which is why computers are so useful: When we use a computer to add a list of ten thousand numbers, it carries out that instruction to the letter, never varying if asked to do it again.

While it's much harder to predict a brain's behavior, it is not, in theory, impossible. Our current inability to do so is simply a byproduct of the brain's vast complexity, and the fact that we don't have a decent manual for it. Goodman wants to fix this.

He's a wiry guy with heavy eyebrows set over dark eyes. His long curly hair, wispy beard, and loose clothing give him an unkempt appearance, but that impression vanishes as soon as he begins to speak. He delivers lectures on the nature of the mind off the top of his head, every word succinct and direct; his mathematician upbringing shines through in his speech.



Goodman left physics for cognitive science following an introspective, minimalist period in his mid-twenties. That a thing has the property of “simplicity,” he thought, has no external reality; it was a label, a judgment, that he was putting on the world. Examining his own motivations, perceptions, and decisions led him to cognitive psychology, a field that endeavors to describe the mind in the same way physics describes gravity, or motion, or time.

That everything in the universe can be described mathematically is the key idea underlying all of Goodman’s research. Newton’s  $F=MA$  describes the motion of a thrown ball, and it works as well for baseballs as it does for planets. Math describes more complicated phenomena, too—such as weather systems and stock markets—but as things get more complex, so too does the math.

Yet if one could track all the myriad variables in those complex systems, physics remains deterministic, born out in the math that describes them. An equation like Newton’s second law of motion, which governs baseballs and planets, will produce the same result every single time, without exception—which is how pitchers can throw strikes and why planets don’t veer off their orbits. Similar equations exist that describe how atoms interact, how electricity flows, how chemicals combine. And since brains are built of the same kinds of atoms and molecules that make up everything else in the universe, there should be math that describes them, too.

“The brain is a physical system,” says Goodman, “and the mind comes out of the brain. So if you believe that you can characterize physics mathematically, then it must be possible to characterize the mind mathematically, even if it might be ridiculously hard.” All of cognitive science is built on the idea that we can understand the mind as a mathematical system. What’s holding us back is our incomplete understanding of the brain; as that understanding increases, so too will our ability to model the brain mathematically. And at the lowest level, all that computers *do* is math—which makes brains and computers a great match.

A physicist studying motion can measure the diameter of a thrown ball, its mass, its density. She can measure how high it flies when thrown, how far, the force required to throw it. All she needs is a stopwatch, a yardstick, and some patience, and she can derive  $F=MA$ , just like Newton did.

Goodman also has to take measurements to devise his math, but he can’t measure brains in the same way that physicists measure balls. Because his math describes behavior—cognitive function—his measurements instead resemble psychological experiments. For example, when studying how the brain learns, he shows a test subject a series of abstract drawings, like Rorschach inkblots. For each drawing, he asks the test subject, “is this a fep or a blop?” (The labels are nonsense words so that the test subject won’t have any prior associations with them.) The first

few times, the subject must make random choices: “sure, that bunch of squiggles looks like a fep. Why not?”

Over the course of many trials, Goodman teaches the subject what a fep and a blop look like: a fep might be a little more angular, a little larger, and blops might tend smaller, with rounded edges. But he never states any of this explicitly, simply correcting wrong guesses and affirming right ones; the subject is allowed to learn, unconsciously, how to identify a fep from a blop.

Think of how you learned to ride a bicycle. You probably can't identify any of the individual corrections you made to balance, or the correct angle to turn the handlebars while leaning around a corner. You learned to balance on the bike unconsciously, a little bit at a time, which is precisely the behavior Goodman seeks to isolate mathematically. By measuring his test subjects' response rates, the number of correct guesses, and repeating the experiment many times, he can start to quantify the cognition behind learning. He can start to build an equation that describes a small part of the mind.

But Goodman is still far from a rigorous model of the brain as a whole. Most cognitive science focuses on bits and pieces, on language or learning or memory or other functions of the brain. Goodman is no exception, and he has had some great successes building the math behind these phenomena: some of his equations can

predict, with a high rate of success, how long a new test subject will take to learn a few from a blob, or into which category she's likely to sort a given drawing.

Though he concentrates on individual pieces of cognition, Goodman is working toward a mathematical framework that describes the mind as a cohesive whole. "Our idea is that there are common computational principles that are used to construct each of those models," he says, "so those are parts of the more general theory," a comprehensive theory—which does not yet exist—of the mind.

Goodman specifically recognizes how narrow modern AI is, and imagines a day when the math he's developing now might be implemented in a computer—for example, a bit of math that describes how we learn language might allow computers to speak to us in conversational English, turning a conversation with the chatbot Eliza from stilted and confusing to natural and free-flowing. She'd pass the Turing Test.

At the bottom of it all, Goodman is interested in finding out what makes humans so flexible. Think of how we train our pets to perform tricks. To teach a puppy to do something as simple as to sit, you need to spend weeks or even months with it, demonstrating the relationship between the command "sit" and the action of sitting, reinforcing with treats the whole time. A human can learn that same command in seconds, no treats required.

Computers are far less flexible. While some programs have been written that can flexibly perform a single task (like recognizing faces), no computer can learn arbitrary tasks as easily as a puppy, much less a human. No computer exists that learns through repetition and demonstration the way humans do; they only know what they've been explicitly "taught" by their programmers. Even machines that appear flexible in one area are only as flexible as they've been programmed to be, and are incapable of performing outside their area of expertise. Tracking down and describing flexibility mathematically is one of Goodman's ultimate goals.

"I think we're pretty far from" a detailed brain-program, says Goodman. "I'm optimistic that we can keep widening the things we can do in one system and introducing more and more flexibility. I assume that we will, someday, be able to [implement a brain in a computer]. I just don't know when someday is."

Building intelligent machines implies some understanding of what it means to be intelligent—which Goodman says is "not a goal for this century." Instead of trying to define something as ambiguous and intangible as intelligence, he says, we should try to pick out certain characteristics of intelligent behavior—like flexibility—and build systems that behave that way.

Goodman's approach is innately descriptive and formalized: he looks at behavior and devises math that expresses that behavior. But others think that such formalism may not be necessary. In fact, says Louise Gunderson, given a choice

between math and biology, she'll pick the biology every time. She and her husband Jim choose to eschew hard math in favor of learning by example—to build an actual machine that exhibits the flexibility Noah Goodman is trying to capture with numbers.

\* \* \*

The Gundersons' lab, Gamma Two Robotics, has a very safe parking garage. When Vigilus, their security-guard robot, is not patrolling the garage against ne'er-do-wells, he can be found inside sometimes sitting quietly in a corner, sometimes trailing wires out behind him, their opposite ends plugged into a laptop or a voltmeter. Vigilus is the Gundersons' latest effort at building robots with humanlike flexibility, robots that can get around complex environments and carry out simple tasks with little to no human supervision. While a far cry from the generalized flexibility of the human brain, Vigilus and his predecessors Basil and Wilma are flexible enough to adapt on the fly to their environment. They gain this flexibility from the programming at their core—a software version of the human brain.

Gamma Two is nestled among the galleries of Denver's art district, but on its walls hang not framed photographs or abstract paintings but technical diagrams and flowcharts, and—most notably—a picture of a brain. A rough approximation of a

brain, actually, with areas labeled “prefrontal cortex” and “cerebellum” and “semantic memory.” The drawing, a chunky diagram all arrows and right angles, is about as far from the convoluted gray meat of a human brain as a camera lens is from an eyeball, but the metaphor is clear: this is how their robots think.

The robots get around in much the same way humans do, by gathering, interpreting, and acting on information about their surroundings. Their sonar and thermal sensors pick up distance and temperature data, which run through a software prefrontal cortex. That cortex—just like a human’s—formulates a goal and makes decisions about how best to achieve that goal. The robots remember and learn, in a limited sense, from experience. They know the difference between a chair and a human (humans being generally warm, chairs less so). They treat those obstacles differently, so when navigating the close quarters of the lab, the robot won’t ask a chair to get out of its way.

“So much of the brain is involved with knowing what’s around you, and the state of your body,” says Louise. The Gundersons built their robots’ brains around that principle, that it is our understanding of our bodies and our environment that allows us to be flexible in achieving goals. In this approach they have taken a decidedly practical tack, eschewing the mathematical formalism of hard cognitive science in favor of what works and the knowledge that they have right now. They

have built a functional approximation of a brain, one that allows their robots to navigate a room, but that lacks humans' mental flexibility.

"From a cognitive psychologist's point of view," says Louise, "this is not a particularly sophisticated model. We basically have a gecko with a really advanced prefrontal lobe. But it's working well enough—it just has to work in the world."

Louise says evolution created the best examples of systems that work in the world. "If you look at the way the brain is laid out, it doesn't change that much" over time, she says; the oldest parts of the brain—the cerebellum, the limbic system—have remained pretty much the same over the course of vertebrate evolution. One of the basic premises of evolution is that if something hangs around that long, it's probably important; if it's not important, you'll evolve away from it. The Gundersons have taken that idea to heart, and programmed simplistic replicas of the oldest parts of the human brain—the parts nature has demonstrated are the most essential.

Those systems include a cerebellum that acts as a switching station, a prefrontal cortex that handles logic, and centers for working memory and episodic memory. As a whole, the Cybernetic Brain allows the robots to maneuver around their environments autonomously and to execute simple instructions—"go to the kitchen and get me a Coke," or "patrol that warehouse for the next 48 hours and let me know if it catches fire."



The Cybernetic Brain only permits the robots to carry out these simple instructions; anything more complicated would require a more complicated brain. But Gamma Two's customers aren't looking for complexity, only service—where Vigilus is a security guard, Basil and Wilma are “home assistance” robots: butlers. The robots are specifically designed to perform simple tasks that might bore humans, or put them in danger.

The Cybernetic Brain looks on paper a lot like a human brain, but the Gundersons recognize that their robots are a far cry from intelligent. It's easy enough to see, watching Vigilus lumber around the lab. He moves slowly and unsurely, pausing for seconds at a time while he tries to identify a new object. He's a little top-heavy, too, which heightens the impression of hesitancy and indecision, like he's never quite sure if he's going to greet you or fall over. But Vigilus is good at taking orders and he's nothing if not attentive—even when asked to sit still, every ten minutes or so he'll blare out, disconcertingly, “I AM STILL HERE AND LISTENING QUIETLY.”

The robots don't learn anything beyond the layout of a given room and a few things about its contents. Vigilus knows how to get from point A to point B. He has routes stored in memory, and he can get to a place he's never been before by stringing those routes together. Over time he figures out that certain routes work more efficiently than others, and will take those better routes more frequently.

This is his uniqueness. His behavior is not hard-coded into him; it comes about as a result of interactions between the parts of his brain. Told to go to the kitchen, Vigilus's prefrontal cortex will do some planning about what's necessary to carry out that order, access his memory to find out what has happened on that route before, and send instructions through his cerebellum to turn his wheels in the correct direction.

"The first step," says Jim, is for the robot to ask "how should I do that?" —a question that exemplifies what the Gundersons call the "subjunctive robot." It's an internal model the robot develops over time, an abstract representation of the world around it, complete with actions and consequences. The robot can navigate this abstract world and make predictions about what might happen in certain circumstances, much like humans can plan out an action without actually doing anything. Jim calls this the "woulds." "So we can ask him what he would do, if we asked him to do something, and he will explain his plan: 'I *would* do this, I *would* do that.'"

"You separate the action from the plan," adds Louise, which is very much the same way humans operate. We avoid destructive things like walking off high ledges or punching our bosses because we store in our brains abstract versions of reality: we understand the concepts of height and gravity and falling and getting fired without having to actually experience any of them. This allows us to manipulate those

abstract concepts virtually, to imagine them in our mind's eye, and to imagine the consequences of what would happen, were we to punch our boss.

Vigilus also maintains an abstract model of his world in memory, and everything in that model has properties—bookcases that don't move, dogs that do, people that might get out of your way if you ask nicely. So when asked what he would do if he were asked to move from the library to the kitchen, he'll remember that the last time he made that trip he had to navigate around some bookshelves, steer around the table, and avoid the dog. He'll also understand that while he'll likely have to deal with the bookcase and the table again, the dog probably won't be lying in the same place this time. What he's doing in imagining this scenario is exactly what humans do: manipulating abstract symbols in his "mind," all without ever setting a wheel in motion.

Jeff Hawkins, author of *On Intelligence*, writes that this is a fundamental characteristic of human intelligence, a principle he calls the "memory-prediction framework of intelligence." Hawkins writes that intelligent creatures gather memories based on experience and build "invariant representations" of those memories—mental symbols that represent *book* or *chair* or *dog* independently of any physical object. We can then make predictions based on those representations about actions taken in the real world.

Of course, it's not difficult to see the limitations of this definition. Vigilus and Basil and Wilma all do this—they build up memories about the world and use those memories to make predictions about what would happen in a given scenario. They even store something akin to invariant representations. One of the first times the Gundersons knew they were on to something promising was when they showed the robot two chairs and it was able to correctly identify a third with no assistance—just based on temperature and a general idea of “chair-ness.”

But to call the robots intelligent is a stretch. They do show surprising flexibility in navigating unfamiliar environments, but they lack the overall flexibility of true intelligence—they couldn't find your lost screwdriver or critique a Picasso. But the Gundersons are more interested in building something *kind of* smart right now more than something truly intelligent later; in fact, they specifically concentrate on fitness rather than intelligence. “[Our] robots don't have to be intelligent; they just have to be will-fitted to their environment,” says Louise—they have to be able to perform a task efficiently and to cope with a reasonable amount of uncertainty. “We'll call it intelligent when it works.”

In order to get their product to market, the Gundersons have built a greatly simplified model of the brain. In doing so they necessarily make their robots less intelligent. In the end they are left with machines that exhibit remarkable flexibility in a certain context, but that lack the generalized intelligence pure researchers seek.

The robots demonstrate just how difficult the problem of strong AI is: when scientists move from theory to practice, somehow intelligence slips through their fingers.

Still, Basil and Wilma and Vigilus work well enough that Gamma Two has customers waiting on a final product, willing to shell out thousands of dollars for a security guard or a butler. They also show that brain-based AI is a viable path forward. For now, however, Gamma Two's customers can't expect HAL 9000 or even Watson or Eliza; they can't expect smarts out of their robots, only service.

\* \* \*

When Vigilus decides to take a path across the room, he examines what he knows to be true about the world, about the room, and about himself. He knows he has a certain amount of battery power left; that the room is forty feet long; and that bookcases need to be circumnavigated, not asked to move. He "imagines" what would happen if he took a longer path or encountered an unexpected obstacle along the way, and what he would do to cope with those imaginary problems.

In other words, Vigilus tells himself a story. It's not an exciting story, sure, but it's a story nonetheless, with a beginning, a middle, and an end; a plot, a protagonist, and a lesson. In telling his story he is doing something fundamentally human—or so says MIT's Patrick Winston, who believes that storytelling is at the very heart of human intelligence.

Winston is a researcher at CSAIL, MIT's Computer Science and Artificial Intelligence Laboratory. CSAIL is on the second floor of the Stata Center, a building that exemplifies modern AI: flash on the outside, boring gray concrete on the inside. The exterior of the building is all surprise corners and curves, strange angles and everything brushed aluminum and bright colors. Those parts of the Stata Center that do come together at right angles usually jut out from between rounded towers or zigzag awnings. From the outside Stata looks like something from Seuss, but its haphazard exterior belies an interior that looks—from the ground floor, anyway—a lot like any other building: a little café, classrooms, offices. Concrete floors.

CSAIL, MIT's largest interdisciplinary lab, is up a flight of stairs and through a set of double-doors. Patrick Winston's office is past a hexapod robot and a few clusters of grubby graduate students, and is much like any other academic's you might find: a coffee maker, a couch, bookshelves. A window at the back looks out onto another of MIT's preeminent labs—Brain and Cognitive Sciences, where teams of scientists work to illuminate the fundamental nature of the mind.

That idea—exploring the mind—links the two buildings across the open space of Vassar Street and through the tempered glass of Patrick Winston's window. On that side of the street, researchers run human test subjects through an MRI scanner in an attempt to determine how the brain is engineered. On this side of the street,

Winston and his colleagues try to engineer systems that act—at least a little—like the human brain.

Many of the scientists at CSAIL belong to the 95 percent of AI researchers who are interested primarily in building better robots or faster search engines. But a few others, like Winston and his colleagues, take a more investigatory tack. Winston is the head of the Genesis Group, researchers “dedicated to the proposition that the time is ripe to answer fundamental questions about the computations that account for human intelligence,” according to the group’s vision statement. Like their Vassar Street neighbors, Genesis is interested in finding out what makes the mind. Instead of neuroscience, though, they start with a simple question.

Intelligence is everywhere, says Winston, in dolphins and dogs and maybe even in ants and trees. But what makes humans different? What is it about our brand of intelligence that makes us *able* to build robots and search engines, to solve complicated problems and to operate so efficiently in such a complex world? This is the question underlying Genesis’ research: not *what is intelligence*—which, as Noah Goodman pointed out, is not a question we’re likely to answer anytime soon—but rather what about *our* intelligence sets us apart?

Winston says that what differentiates us from our primate cousins is that we’re incurable storytellers. Just like *Vigilus*, we tell ourselves stories in order to make predictions about how best to get around in the world. What happens if you go

out in the cold with no jacket? You get cold, so don't leave the fleece behind. We don't need to go outside to know this will happen, because it's clear in the story. We also link stories together to tell different stories. Going outside bundled up will keep you warm, but will that jacket fit in your locker once you get to the gym? We draw on our experience to build these stories and act them out in our minds, and much of what we understand as common sense is derived from using our senses to experience events both real and imagined.

To illustrate this, Winston tells the story of having a friend over to his house to install a new table saw. While they were working, the friend advised Winston that he should never wear gloves while operating the saw. "At first I was mystified," says Winston, "then it occurred to me that a glove could get caught in the blade. No further explanation was needed because I could imagine what would follow." He took the small story from his friend—you shouldn't wear gloves when you operate table saws—and combined it with his perceptions, his knowledge of gloves and spinning blades, to tell a new story inside his own mind. "I told myself a story about something I have never witnessed, and [now] I will have the common sense to never wear gloves when I operate a table saw."

To investigate this story-based understanding of intelligence, Genesis designed a piece of software that can "understand" what stories are *about*—not necessarily the words themselves, but their underlying themes or ideas, much as



Winston's table-saw story wasn't about table saws so much as it was about *danger*.

The Genesis software currently knows about two dozen stories, including the following simplistic synopsis of Shakespeare's Macbeth:

Macbeth, Macduff, Lady Macbeth, and Duncan are persons. Macbeth is a thane and Macduff is a thane. Lady Macbeth, who is Macbeth's wife, is greedy. Duncan, who is Macduff's friend, is the king, and Macbeth is Duncan's successor. Macbeth defeated a rebel. Witches had visions and talked with Macbeth. The witches made predictions. Duncan became happy because Macbeth defeated the rebel. Duncan rewarded Macbeth because Duncan became happy. Lady Macbeth, who is Macbeth's wife, wants to become the queen. Lady Macbeth persuades Macbeth to want to become the king. Macbeth murders Duncan. Then, Lady Macbeth kills herself. Dunsinane is a castle and Burnham Wood is a forest. Burnham Wood came to Dunsinane. Macduff had unusual birth. Macduff fights with Macbeth and kills him. The predictions came true.

The system also contains the following description of a Russian cyberattack on Estonia's network infrastructure:

Estonia and Russia are countries. Computer networks are artifacts. Estonia insulted Russia because Estonia relocated a war memorial. Someone attacked Estonia's computer networks after Estonia insulted Russia. The attack on Estonia's computer networks included the jamming of web sites. The jamming of web sites showed that someone did not respect Estonia. Estonia created a center to study computer security. Estonia believed other states would support the center.

Given these two stories, a rudimentary understanding of English grammar, and some commonsense rules (like 'if X kills Y, then Y becomes dead'), the Genesis software is able to determine that revenge is involved in both stories despite the stories' vastly different content and even though neither story uses the word "revenge."

While the software cannot yet combine two stories to make a third—a key component of human intelligence, Winston believes—it can tailor its own storytelling to match the needs and desires of a listener; in some sense it is smart enough to pick out what is necessary to relate and what isn't. "If the system [thinks] you're from Mars and you don't know anything," says Winston, "then it will say 'Macbeth murdered Duncan; Duncan is dead'—because you don't know what murder does. If it thinks you are smarter, it will leave that out."

The fact that we humans tailor our stories to our audience shows that we're telling *ourselves* stories, unconsciously, about our audience's points of view and motivations—we're imagining what would happen if we were to use different words to communicate, and making judgment calls about which words will be most effective. These commonsense reactions to our audience are one of the most basic ways we communicate with each other. Teaching computers the same concepts is an important step toward making them more relatable, and more conversationally flexible.

Much of Genesis's research is predicated on the idea that "we don't really understand something unless we can build it," says Winston; "it's an engineering approach to a scientific question." The idea is that if the software can reproduce human-like behavior, at least they've found a way to model that behavior, even if it's

not precisely the way humans think. If it fails, says Winston, at least “we’ve got a new problem to work on.”

The Genesis scientists’ ultimate purpose is to understand human understanding; their software is specifically not an attempt to simulate intelligence, but to learn how we think, how we combine ideas to tell stories, and the role stories play in human interaction and learning. The Genesis software won’t have any application beyond pure research, but the research may hint at something fundamental.

Genesis is an approach to AI that emphasizes questions about what makes humans different. Their software is successful not because it is intelligent—it’s not even close—but because it *points* at intelligence. Later generations of researchers might use math or computer algorithms to implement the behavior that Winston and his colleagues unearth, but Winston believes math and algorithms are simply tools that will be used to solve the larger problems. “If you start off with the tool rather than the problem,” he says, “you’re not going to answer the questions that are important. You have to figure out what the problem is before you can bring the tool to it.”

\* \* \*

At first blush, you might not guess Ben Goertzel is a scientist. He speaks in surfer tones, slow and methodical, with a soft lisp like he's perpetually tipsy. He wears his curly hair long and gesticulates wildly with spindly limbs while he talks, his round John-Lennon glasses casting circular shadows on his craggy skin. He writes avant-garde fiction and composes rambling piano music in his spare time. He has been known to appear in public in a zebra-print cowboy hat.

Impressions can be deceiving, though. Goertzel earned a Bachelor's degree at age 19, and a PhD in mathematics at 23. He's founded four companies and an AI conference series, advises a number of research efforts, teaches, lectures, and believes that one day, artificial intelligence will take over the Earth—probably with direct assistance from him.

His approach to building strong AI reflects his eclectic personality. He has scrutinized entrepreneurs like the Gundersons who mimic the brain's organization in their robots; bioengineers like Poon who appreciate the elegance of self-organization; mathematicians like Goodman who find promise in numbers and theory; and AI researchers like Winston who hunt for clues about what sets humans apart. Goertzel sees all these varied approaches and concludes that none of them alone is likely to pave the path to strong AI—but all of them, together, just might.

So he picks and chooses from cognitive science and mathematics, takes cues from human behavior, and ties it all up in a computer-science bow. But his most

fundamental observation about the best way to build AI is that humans, for all our faults, are good learners. We do it flexibly and quickly, and we do much of our best learning when we're young. So Goertzel's AI projects are not designed to be intelligent right away. Instead, he focuses on building systems that learn in ways similar to how humans learn. His AI programs—his "agents"—start as infants, with many of the same fundamental cognitive capacities that infant humans develop.

Babies are sponges for information, taking in sensory impressions of their world, processing and categorizing. Eventually they start to recognize objects, from their parents' faces, to spoons, to jars of pureed yams. They learn "object permanence:" that a spoon full of yam doesn't disappear from the universe just because it disappears from their field of view. They start to recognize objects' shapes, and this familiarity allows them to learn the rules of cause and effect. Parents carry spoons; spoons deliver food; food cures hunger. Cry when you're hungry and a parent appears, spoon in hand. Yams ensue. These leaps of logic are the foundation for basic problem-solving skills.

Goertzel's agents learn in much the same way, and have many of the same motivations as humans: they get tired, they get afraid, they get hungry. With basic cognitive processes in place, they can adapt to their environment in order to achieve goals that satisfy their basic needs—just as a hungry toddler might open a series of kitchen drawers, using them as steps to reach the cookie jar.

But the kitchens of Goertzel's agents look a little unusual. Instead of being made of marble and wood and stainless steel, their kitchens—in fact, their entire world—is made of something resembling Legos. The agents live in a virtual world inside computers at Goertzel's research laboratories in Hong Kong—a virtual world composed of tens of thousands of blocks. Blocks for kitchens, blocks for trees. Blocks for rivers to cross and walls to build, blocks for mountains to scale. A sea of blocks, with unlimited potential for problems, and their solutions.

Goertzel designed his block universe because it simplifies the agents' interactions with their environment—it's easier to build things out of Legos than to build them out of wood or stone. The block universe also has huge creative potential: just as you can build a car out of Legos, or a dragon, or a house, the agents can use the blocks to build anything they need.

So each agent has basic reasoning skills, motivations such as hunger and fear, and lives in a world that facilitates problem solving. It's a perfect environment for learning, which is precisely what Goertzel wants his agents to do. Even if they can't be called "intelligent" quite yet, his agents display some startlingly adaptive behavior. One hungry agent spotted some "food" out of reach on a high shelf, and just like the toddler and the kitchen drawers, built a block staircase in order to reach his meal. Another agent, sensing the presence of a hostile agent out to steal its food, built a wall to hide behind.

This is the kind of simple problem-solving that human toddlers might display, which implies Goertzel's learning algorithms are doing something right. After going through the formative periods of infancy and toddlerhood, Goertzel hopes his agents will display more complex problem-solving behavior as "children," develop a more sophisticated picture of the world as "young adults," and eventually start to display something approaching real intelligence. Ultimately, says Goertzel, he hopes to grow an intelligence "way beyond human level."

While his agents have a thing or two in common with humans developmentally, at lower levels the comparison falls apart. Goertzel models his agents' high-level cognition on current theories in cognitive psychology: they have several kinds of memory (such as *episodic*, memories about events in one's life; or *procedural*, memories about how to do things), the ability to reason, to deduce, and to infer. But how each of those things is actually carried out inside an agent's "brain" differs vastly from how it likely happens in the human brain. "We're trying to make a smart system," he says. "If there are ways to exploit algorithms or computer hardware that make the system smart, but may make it deviate from the ways humans do things, we're always willing to do that."

What is more, Goertzel says that if his agents are going to evolve past human-level intelligence, using computer-science methods to get them there is critical. Evolution is a messy business, and millions of years of it has resulted in a system that

is enormously flexible, enormously powerful, and enormously inefficient. “The brain goes through a lot of complicated gymnastics to do some really simple stuff,” he says. Recognizing a face or recalling a memory or making a decision involves hundreds of millions of neurons all firing at each other, chemicals coursing through the brain, electricity zapping like lightning storms across synapses. A few lines of code, says Goertzel, could do any of those things far more efficiently. In essence, he wants to streamline evolution and fashion a unique, computer-based mind.

So while he builds on a framework of human cognition and is inspired by developmental psychology, Goertzel is careful to leave out the parts of the human brain that might get in the way of strong AI. For example, in 1956, Harvard psychologist George Miller famously suggested that humans can only store about seven things in short-term memory. If this is true, it fundamentally—and arbitrarily—limits humans’ ability to process information and, arguably, limits our intelligence. “Why would you want to replicate that in a computer?” asks Goertzel. “If the goal is to do things well, why not let your computer have indefinite short-term memory?”

By learning from the things the brain does well and improving on its weaknesses with computer science methods, Goertzel thinks he is finding a handle on what it will take to build strong AI. But while this goal has occupied much of his



personal and professional life for the last quarter-century, he still thinks of it as a means to an end.

Goertzel believes that engineering brain-like systems without the weaknesses of brains will ultimately result in a brand of intelligence that exceeds our own. Such an intelligence would be difficult to predict, much less control, but as Goertzel puts it, “that’s the risk you take” when toying with big ideas. Nonetheless, he is optimistic about the role strong AI will play in the future of humankind. A powerful AI without the limitations of the human brain could contain a much larger “problem space” than a human can, taking into account a massive number of variables, from global economics to intricate socio-political conditions. A strong AI in charge of its own research lab could be put on the task of curing cancer, or designing efficient energy sources.

More to the point, Goertzel believes that strong AI would be capable of solving virtually any problem we give it, which means it’s a kind of philosophical catch-all: he’s working toward solving all the world’s problems at once. Inventing a tool that can do anything, he believes, is *the* most important problem.

In Goertzel’s view, there is nothing fundamental standing in his way—just time and money. With enough patience and a sufficiently detailed roadmap, he suggests, strong AI is inevitable. “Of all the really big amazing things to do,” he says, “what is appealing about AI is that you can just do it by sitting down and writing

code.” His strong-AI children, meanwhile, live and grow and learn in their blocky universe, building and discovering, and, just perhaps, biding their time.

\* \* \*

You wake on a summer morning in your darkened bedroom, the only light coming from a narrow gap between the drapes. You’re half asleep. You stretch a little and push the sheets down, rub the sleep from your eyes. Then, creeping up from the back of your mind, you remember you have a meeting at nine o’clock. The faintly recalled memory jolts your mind awake.

You understand several things in that first flash of full consciousness. Your car is in the shop, so you’ll have to ride your bike. You know it’s summertime, and the tone of the light seeping in between the curtains tells you it’s already full sun outside; it’s going to be a hot day. Traffic on South Street will be backed up this time of morning, but the bike should help you through that—at the same time, you’ll have to hustle to maneuver through the gridlock, and you’re likely to be sweaty by the time you get to the office. You’ll need to bring a change of clothes.

This cascade of thought happens in a heartbeat—you’re still rubbing your eyes. Your mind snaps from *I have an appointment* to *I’ll need a change of clothes* with barely a moment between. Your brain has blended memory, experience, abstraction,

and forethought into a seamless whole, making it possible for to solve a complicated problem with barely any effort.

Imagine how a computer might solve this same problem. First, it needs its goals explicitly stated—get to work on time, and don't make a bad impression when you get there. It needs all of the variables defined—car in the shop, bike available but slower, traffic conditions, weather and temperature, distance, and so on. Then it needs to evaluate all of the second-degree variables, the possible outcomes of the first variables—time to reach the office, effort required to do so, how the body might react in those conditions, whether it's socially acceptable to be drenched in sweat at the office.

Finally, it needs to evaluate all possible solutions. It could, for example, invent teleportation and zap instantaneously to the office, but that might take a lot of time, and there's that nine o'clock appointment to consider. Maybe it could quit the job, which has its own set of ramifications. It could estimate the level of discontent among your officemates were it simply to disregard social niceties and show up for the meeting dripping with sweat. There are an infinite number of solutions to this problem—an infinite number of stories—some of which make sense (take the bus), some of which don't (move to Pluto). The human brain is able to tell the viable solutions from the absurd, and discard the latter without even consciously considering them.

The computer, though, has to perform each step of the process deliberately, and one at a time. If some variable changes—it's not as hot outside as the light makes it seem, the bicycle has a flat tire—it needs to reevaluate all of the variables in the new context; it has to solve the problem all over again, running its program, its instruction set, from start to finish, just like any other computer.

The beauty and mystery and power of the human brain lies in its unique capacity to take in discrete pieces of information—the tone of light on the ceiling, the recollection of a leaky fuel line on the Chevy—and pinpoint the pieces relevant to the problem at hand. It distills information, synthesizes solutions, and does it all automatically, with barely any conscious intervention by the user, the carrier, the human.

It's no wonder we've had such a hard time replicating such an elegant problem-solving machine. The brain is the most intricate and complex structure in the known universe, and we still have no idea how it is engineered. So we look at how we behave—the second-order variable—and make inferences about what must be going on behind the scenes. We take MRI pictures, we run experiments. We explore the brain.

The cycle of promise and failure that has so reliably characterized AI for the last sixty years is simply a byproduct of this as-yet incomplete exploration of the brain. AI has to make guesses, just like any other fledgling science: Ptolemy

speculated that the universe was a set of nested spheres with Earth at their center until Copernicus modified that model to put the Sun at the center of the solar system. As our understanding of the brain and the mind increases, AI will progress from faulty simulacra like Eliza and narrow, problem-specific programs such as Deep Blue or Google to adaptable, strong AI that can both compete with and bolster our own intelligence. It will be a synthesis of mind and technology that will lead AI out of the cycle of the last half-century.

But it won't happen all at once, and it won't come from just one place. Gamma Two's robots show us that even a rudimentary model of the brain can help programs navigate the world and accomplish simple tasks. Noah Goodman's formalized mathematical models promise to define the mind as precisely as physics describes black holes. Chi-Sang Poon's work to emulate the physical structures of the brain may one day serve as the substrate on which we build a mind.

But no single approach holds the answer. Just as Copernicus and Galileo improved Ptolemy's model of the heavens, future generations of scientists—biologists and engineers and mathematicians and entrepreneurs—will build on the work of Goodman and Poon and Winston and many others. And it will happen a little at a time; it might even sneak up on us. "What we're going to see is not, 'we can't, we can't, we can't, then *boom*, we can,'" says Goodman. "It's going to be that the scope of things we can do in a single system grows and becomes more flexible

and more human-like incrementally. At some point we'll wake up and realize, 'hey, this [computer program] can do most of the things that *we* can do.'

Computer intelligence will be the artifact, the physical manifestation of this many-faceted research, and it may well conform to the creative musings of those, like Goertzel, who think it will solve all the world's problems. But such an intelligence, by definition, will be autonomous and sovereign as any human, and equally as subject to whim and folly. Some futurists speculate that the creation of a super-intelligent AI will herald the coming of the Singularity, when humans surrender (willingly or not) their position as the dominant beings on the planet and past which it becomes impossible to predict the course of human events.

But this is wild speculation, and most researchers take a more grounded approach. Some believe that a super-intelligent computer is inevitable, but that it lies so far in the future—centuries or millennia—as to be of only academic concern to us today. Scientists with their feet on the ground recognize that synthesizing intelligence in a machine is not the ultimate goal but a byproduct of a loftier pursuit: to know the mind. When we learn what makes us intelligent, when we understand why we understand, we will be able to grow truly without limit.

## Endnotes

4. *Vigilus is undergoing*: Gheysar.
5. *Vigilus has a brain*: Gundersons.
5. *More than sixty years ago*: McCarthy, John, et al. "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence."
6. *Concerted efforts began*: Ibid.
7. *"the science and engineering of making intelligent machines"*: McCarthy, John. "What is Artificial Intelligence?"
7. *he proposed an experiment*: Turing.
8. *programmers gather for the yearly Loebner Prize*: Loebner Prize.
9. *Eliza was the first "chatbot"*: Weizenbaum.
10. *When researchers at Cornell*: Cornell Creative Machines Lab. *AI vs. AI*.
11. *proposed a thought experiment*: Searle. "Minds, brains, and programs."
12. *"the coming days of ascendant computers"*: McFadden.
13. *The same tactic was visible*: IBM. "IBM–Watson."
13. *a peek behind the scenes*: Ferrucci.
13. *"Whenever we understand..."*: Winston, 22 Sep 2011.
14. *"Strong AI"*: Searle. "Minds, brains, and programs."
14. *"in exactly the same sense..."*: Ibid.
14. *doing original scientific research*: Goertzel, 9 Nov 2011.
14. *The most ambitious*: Ibid.
16. *Wernicke's Area*: Bogen.
18. *a point he makes emphatically*: Poon, 19 Jan 2012.
19. *some hundred billion of the former*: Williams.
20. *Many other attempts*: EPFL.
21. *"algorithms to explain the mind"*: Poon, 19 Jan 2012.
21. *Poon is the first to reproduce it*: Ibid.
22. *"We already know how neurons work"*: Ibid.
25. *Goodman left physics*: Goodman, 22 Mar 2012.
26. *"The brain is a physical system"*: Goodman, 6 Jan 2012.
26. *he shows a test subject*: Goodman, 22 Mar 2012.
28. *"there are common computational principles"*: Goodman, 6 Jan 2012.
29. *"I think we're pretty far"*: Ibid.
30. *she'll pick the biology*: Gunderson, Louise.
31. *"So much of the brain"*: Ibid.
32. *"From a cognitive psychologist's point of view"*: Ibid.
32. *"If you look at the way"*: Ibid.
32. *Those systems include a cerebellum*: Johnson.

33. "I AM STILL HERE": Gundersons.
34. "how should I do that?": Gunderson, Jim.
34. "You separate the action": Gunderson, Louise.
35. "memory-prediction framework": Hawkins, 104.
35. "invariant representations": Hawkins, 75.
36. *One of the first times*: Gunderson, Jim.
36. "robots don't have to be intelligent": Gunderson, Louise.
39. *95 percent*: Winston, 30 Nov 2011.
39. "dedicated to the proposition": Winston, Genesis home page.
39. *differentiates us from our primate cousins*: Winston, "The Strong Story Hypothesis"
40. *Winston tells the story*: Ibid.
41. *about two dozen stories*: Winston, 11 Jan 2012.
42. *a key component of human intelligence*: Winston, 22 Sep 2011.
42. "If the system [thinks] you're from Mars": Ibid.
42. "we don't really understand something": Ibid.
43. *specifically not an attempt*: Winston, 11 Jan 2012.
43. "If you start off with the tool": Winston, 22 Sep 2011.
44. *avant-garde fiction...cowboy hat*: Goertzel, Ben Goertzel's Website.
44. *direct assistance from him*: Ibid.
45. "object permanence": Goertzel, 10 Apr 2012.
46. *made of something resembling Legos*: Ibid.
46. *agent spotted some "food"*: Ibid.
47. "way beyond human level": Goertzel, 9 Nov 2011.
47. "We're trying to make a smart system": Ibid.
47. *if his agents are going to evolve*: Ibid.
48. "a lot of complicated gymnastics": Ibid.
48. *humans can only store about seven things*: Miller.
48. "Why would you want to replicate that": Goertzel, 10 Apr 2012.
49. "that's the risk you take": Ibid.
49. *working toward solving all the world's problems*: Ibid.
49. "Of all the really big amazing things": Ibid.
53. "What we're going to see": Goodman, 6 Jan 2012.
54. *Some futurists speculate*: Kurzweil, 2005.
54. *super-intelligent computer is inevitable*: Poon, 19 Jan 2012.



## References

- Allen, Frederick. "Sorry, Jeopardy Watchers: No Computer Will Ever Be Remotely Human." *Forbes*, 15 Feb 2011. Web. 5 Jan 2012.
- Aur, Dorian. "From Neuroelectrodynamics to Thinking Machines." *Cognitive Computation*. (2011): 1-9. Web. 26 Jan. 2012. <<http://dx.doi.org/10.1007/s12559-011-9106-3>>.
- Best, John. *Cognitive Psychology*. 5th ed. Belmont, CA: Wadsworth Publishing Company, 1999. Print.
- Bogen, J.E., and G.M. Bogen. "Wernicke's Region—Where Is It?" *Annals of the New York Academy of Sciences*. 280. (1976): 834-843. Print.
- Carpenter, Rollo. "Cleverbot." *Cleverbot.com - A Clever Bot*. N.p., 2011. Web. 15 Oct 2011. <<http://cleverbot.com/>>.
- Cornell Creative Machines Lab. *AI vs. AI. Two chatbots talking to each other*. 2011. Video. YouTube. Web. 15 Oct 2011. <<http://www.youtube.com/watch?v=WnzlbyTZsQY>>.
- Crevier, Daniel. *AI: The Tumultuous History of the Search for Artificial Intelligence*. New York: HarperCollins, 1993. Print.
- Ecole Polytechnique Fédérale de Lausanne. "EPFL | Bluebrain." EPFL, 2011. Web. 25 Jan 2012. <<http://bluebrain.epfl.ch/>>
- Ferrucci, David. "The AI Behind Watson: The Technical Article." *AI Magazine*. (2010): n. page. Web. 5 Jan 2012. <<http://www.aaai.org/Magazine/Watson/watson.php>>.
- Franklin, Stan et al. "LIDA: A Computational Model of Global Workspace Theory and Developmental Learning." *Journal of Comparative Neurology and Psychology* (2007): n. page. Print.
- Fritz, Sandy, comp. *Understanding Artificial Intelligence*. New York: Warner Books, Inc, 2002. Print.
- Gheysar, Pamela. "Robot Completes Full Day's Work, Autonomously." *Gamma Two News*. Gamma Two Robotics, 29 Jan 2012. Web. 18 Feb 2012. <<http://www.gamma-two.com/vigulus-full-shift-012912.pdf>>.
- Goertzel, Ben. "Cognitive synergy: A universal principle for feasible general intelligence." *8th IEEE International Conference on Cognitive Informatics*. (2009): 464-468.

Web. 26 Jan. 2012.

<<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5250694&isnumber=5250668>>.

Goertzel, Ben. "Ben Goertzel's Website." N.p., 4 Apr 2012. Web. 17 Sep 2011.

<<http://wp.goertzel.org/>>.

Goertzel, Ben. "The Mind-World Correspondence Principle: Toward a General Theory of General Intelligence." *Dynamical Psychology 2011*. 17 Sep 2011. Web. 12 Jan 2012. <[http://dynapsyc.org/2011/Mind\\_World\\_Correspondence.pdf](http://dynapsyc.org/2011/Mind_World_Correspondence.pdf)>.

Goertzel, Ben. "Toward a Formal Characterization of Real-World General Intelligence." *Proceedings of the Third Conference on Artificial General Intelligence*. (2010): n. page. Web. 21 Dec. 2011. <[http://agi-conf.org/2010/wp-content/uploads/2009/06/paper\\_14.pdf](http://agi-conf.org/2010/wp-content/uploads/2009/06/paper_14.pdf)>.

Goertzel, Ben, et al. "Cognitive Synergy between Procedural and Declarative Learning in the Control of Animated and Robotic Agents Using the OpenCogPrime AGI Architecture." *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*. (2011): 1436-1441. Web. 21 Dec. 2011.

<<http://www.aaai.org/ocs/index.php/AAAI/AAAI11/paper/view/3562/4096>>.

Goodman, Noah, Joshua Tenenbaum, et al. "A Rational Analysis of Rule-Based Concept Learning." *Cognitive Science*. 32.1 (2008): 108-154. Print.

Goodman, Noah, et al. "Intuitive Theories of Mind: A Rational Approach to False Belief." *Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society*. (2006): n. page. Web. 13 Jan. 2012. <<http://web.mit.edu/cocosci/Papers/pos785-goodman.pdf>>.

Hawkins, Jeff. *On Intelligence*. 1st ed. New York: St. Martin's Press, 2004. Print.

"Home Page of the Loebner Prize in Artificial Intelligence." *The Loebner Prize*. Loebner Prize, 11 Jan 2011. Web. 11 Feb 2012. <<http://www.loebner.net/Prizef/loebner-prize.html>>.

IBM. "IBM Research: Deep Blue." IBM, 1997. Web. 3 Jan 2012. <<http://www.research.ibm.com/deepblue/meet/html/d.3.shtml>>.

IBM. "IBM-Watson." IBM, 2011. Web. 3 Jan 2012. <<http://www-03.ibm.com/innovation/us/watson/index.html>>.

- Johnson, Keith, and J. Alex Becker. "The Whole Brain Atlas." Harvard University, 1999. Web. 28 Oct 2011. <<http://www.med.harvard.edu/aanlib/home.html>>.
- Kurzweil, Ray. *The Age of Intelligent Machines*. Cambridge, Mass.: The MIT Press, 1990. Print.
- Kurzweil, Ray. *The Age of Spiritual Machines*. New York: Penguin Putnam Inc, 1999. Print.
- Kurzweil, Ray. *The Singularity is Near*. New York: Penguin Group, 2005. Print.
- Markoff, John. "Computer Wins on Jeopardy!: Trivial, It's Not." *New York Times* 16 Feb 2011, n. pag. Web. 23 Feb 2011. <[http://www.nytimes.com/2011/02/17/science/17jeopardy-watson.html?\\_r=1&pagewanted=all](http://www.nytimes.com/2011/02/17/science/17jeopardy-watson.html?_r=1&pagewanted=all)>.
- McCarthy, John. "What is Artificial Intelligence?" Stanford University, 12 Nov. 2007. Web. 19 Jan 2012. <<http://www-formal.stanford.edu/jmc/whatisai/>>
- McCarthy, John, et al. "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence." 31 Aug 1955. Web. 4 Oct 2011. <<http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>>.
- McFadden, Robert. "Computer in the News: Kasparov's Inscrutable Conqueror." *New York Times* 12 May 1997, n. page. Web. 14 Jan. 2012. <<http://www.nytimes.com/library/cyber/week/051297blue.html>>.
- Miller, George. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." *Psychological Review*. 101.2 (1955): 343-352. Print.
- Minsky, Marvin. "Steps toward Artificial Intelligence." *Proceedings of the Institute of Radio Engineers*. 49.1 (1961): 8-30. Print.
- Minsky, Marvin. "Why People Think Computers Can't." *AI Magazine*. 3.4 (1982): 3-15. Print.
- Minsky, Marvin, ed. *Semantic Information Processing*. Cambridge, Mass.: The MIT Press, 1968. Print.
- Olson, Sander. "Interview of Noah Goodman of the Cognitive Science Group at MIT." *Next Big Future*. 13 Apr 2010. Web. 22 Feb 2012. <<http://nextbigfuture.com/2010/04/interview-of-noah-goodman-of-cogitive.html>>.

Pinker, Steven. *The Language Instinct: How the Mind Creates Language*. 1st ed. New York: William Morrow and Co., 1994. Print.

Poon, Chi-Sang, and Kuan Zhou. "Neuromorphic silicon neurons and large-scale neural networks: challenges and opportunities." *Frontiers in Neuromorphic Engineering*. 5.0 (2011): n. page. Web. 3 Jan. 2012.  
<[http://www.frontiersin.org/Journal/FullText.aspx?s=755&name=neuromorphic\\_engineering&ART\\_DOI=10.3389/fnins.2011.00108](http://www.frontiersin.org/Journal/FullText.aspx?s=755&name=neuromorphic_engineering&ART_DOI=10.3389/fnins.2011.00108)>.

Rachmuth, Guy, Chi-Sang Poon, et al. "A biophysically-based neuromorphic model of spike rate- and timing-dependent plasticity." *Proceedings of the National Academy of Sciences*. 108.49 (2011): E1266-E1274. Print.

Rees, Geraint, Gabriel Kreiman, and Cristof Koch. "Neural correlates of consciousness in humans." *Nature Reviews Neuroscience*. 3. (2002): 261-270. Print.

Searle, John. *Mind, Language and Society*. 1st ed. New York: Basic Books, 1999. Print.

Searle, John. "Minds, brains, and programs." *Behavioral and Brain Sciences*. 3.3 (1980): 417-424. Print.

Simon, Herbert. *Models of My Life*. Cambridge, Mass.: The MIT Press, 1996. Print.

Steyvers, Mark, and Joshua Tenenbaum. "The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth." *Cognitive Science*. 29.1 (2005): 41-78. Print.

Stillings, Neil, et al. *Cognitive Science: An Introduction*. 2nd ed. Cambridge, Mass.: The MIT Press, 1995. Print.

Tenenbaum, Joshua, Noah Goodman, et al. "How to Grow a Mind: Statistics, Structure, and Abstraction." *Science*. 331.6022 (2011): 1279-1285. Print.

Tononi, Giulio, and Cristof Koch. "The Neural Correlates of Consciousness: An Update." *Annals of the New York Academy of Sciences*. 1124. (2008): 239-261. Print.

Turing, Alan. "Computing machinery and intelligence." *Mind*. 59. (1950): 433-460. Print.

Vul, Edward, et al. "One and Done? Optimal Decisions From Very Few Samples." *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*. (2009): 66-72. Print.

Weizenbaum, Joseph. "ELIZA — A computer program for the study of natural language communication between man and machine." *Communications of the ACM*. 9.1 (1966): 36-45. Print.

Williams, Robert, and Karl Herrup. "The Control of Neuron Number." *Annual Review of Neuroscience*. 11. (1988): 423-453. Web. 19 Oct 2011.

Winston, Patrick. "Genesis Home Page." *Genesis*. MIT, 21 Apr 2012. Web. 21 Sep 2011. <<http://groups.csail.mit.edu/genesis/>>.

Winston, Patrick. *Learning Structural Descriptions from Examples*. Massachusetts Institute of Technology, 1970. Web. <<http://hdl.handle.net/1721.1/13800>>

Winston, Patrick. "The Strong Story Hypothesis and the Directed Perception Hypothesis." *AAAI Fall Symposium Series*. (2011): n. page. Web. 15 Dec. 2011. <<http://www.aaai.org/ocs/index.php/FSS/FSS11/paper/view/4125/4534>>.

## **Interviews**

Goertzel, Ben. Research Professor, Fujian Key Laboratory for Brain-like Intelligent Systems, Xiamen University. Telephone interview. 9 Nov 2011.

Goertzel, Ben. Research Professor, Fujian Key Laboratory for Brain-like Intelligent Systems, Xiamen University. Skype interview. 10 Apr 2012

Goodman, Noah. Assistant Professor of Psychology and Computer Science, Stanford University. Skype interview. 6 Jan 2012.

Goodman, Noah. Assistant Professor of Psychology and Computer Science, Stanford University. Skype interview. 22 Mar 2012.

Gunderson, Jim. CTO, Gamma Two Robotics. Personal interview. 29 Dec 2011.

Gunderson, Louise. CEO, Gamma Two Robotics. Personal interview. 29 Dec 2011.

Harrell, Fox. Associate Professor of Computer Science, Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology. Personal interview. 26 Oct 2011.

Koch, Christof. Professor of Biology and Engineering, Koch Lab, California Institute of Technology. Personal interview. 16 Oct 2011.

Poon, Chi-Sang. Principal Research Scientist, Harvard-MIT Health Sciences and Technology, Massachusetts Institute of Technology. Personal interview. 19 Jan 2012.

Poon, Chi-Sang. Principal Research Scientist, Harvard-MIT Health Sciences and Technology, Massachusetts Institute of Technology. Personal interview. 7 Mar 2012.

Szolovits, Peter. Professor of Computer Science, Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology. Personal interview. 19 Sep 2011.

Winston, Patrick. Professor of Computer Science, Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology. Personal interview. 22 Sep 2011.

Winston, Patrick. Professor of Computer Science, Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology. Personal interview. 30 Nov 2011.

Winston, Patrick. Professor of Computer Science, Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology. Personal interview. 11 Jan 2012.

Wissner-Gross, Alex. Institute Fellow, Harvard University Institute for Applied Computational Science. Personal interview. 16 Oct 2011.