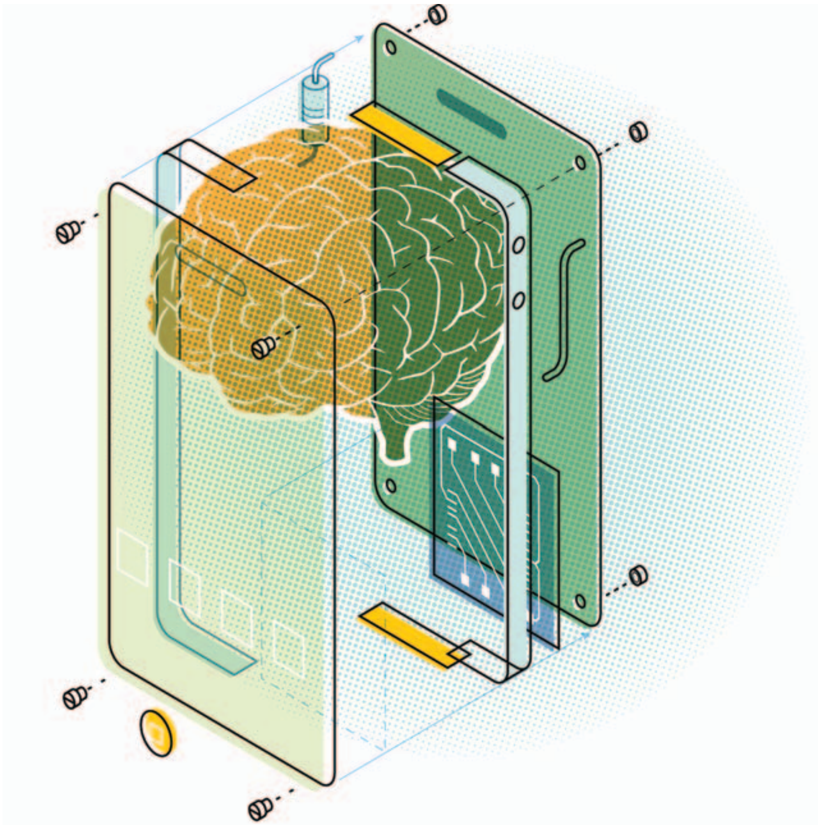


MIND ARCHITECTURE

Does My Smartphone Really Love Me?

The movie Her *makes a compelling case for a computer program with feelings. Is that actually possible?*

> *Samantha: "And then I had this terrible thought. Are these feelings even real? Or are they just programming?"*

An anodyne and restrained Theodore Twombly falls in love with Samantha, the female persona of his computer operating system, in the recent movie *Her.* The romance begins as they overcome their somewhat awkward initial encounter and settle into an easygoing relationship. She arranges his life and tries to fix him up with a date. He tells her about his dreams. They banter, bond over his ongoing divorce and have endless conversations about people, events and desires as they explore the Los Angeles of the near future. Little by little, she reels him in.

BY CHRISTOF KOCH

Christof Koch is chief scientific officer at the Allen Institute for Brain Science in Seattle. He serves on *Scientific American Mind's* board of advisers.

Send suggestions for column topics to editors@SciAmMind.com

Theodore's face lights up every time she calls. He is clearly in love with this ethereal being. He panics when he can't talk to her. The honeymoon ends after a disastrous attempt at a ménage à trois. Their deteriorating relationship is finally rent asunder after she confesses that she is carrying on intimate conversations with hundreds of others.

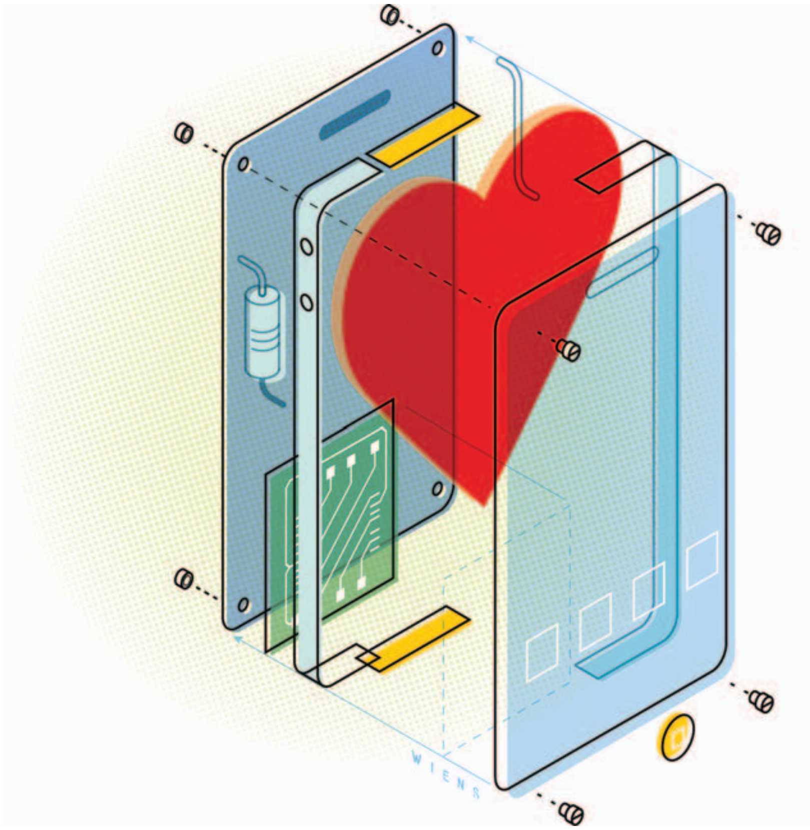
What is remarkable about this smart movie is that the story is so convincing. It leads you to believe that a man can fall in love with a disembodied female voice, an applet. While Theodore is flesh and blood, Samantha is a software program. She is linked to him by his omnipresent phone, a kind of glorified Siri living in the cloud, where she is simultaneously interacting with thousands of other people.

Her does not delve into the profound questions that the existence of a Samantha implies: Can she really love Theodore, or is she just feigning it? Does it feel like anything to be her? Indeed, can a software construct ever be conscious, or is it condemned to a zombie existence, cleverly programmed to respond appropriately but ultimately without any feelings whatsoever?

Until the introduction of smartphones, most people would have dismissed a Samantha-like being as implausible. The smartphone is a product of the relentless progress in the computer industry. The emergence of ever more powerful machines replete with ever deeper memory banks is being driven by a hypercompetitive marketplace and by Moore's law, the empirical observation in the semiconductor industry that the number of transistors on a chip doubles every two years or so.

Availability of better hardware has meant that machine-learning techniques invented several decades ago can finally be put to work in software algorithms to allow computers and robots to perform tasks that all of us carry out automatically, day in and day out, without a second thought. These tasks have typically been some of the most difficult ones for computers to execute, but real progress is now being made. Machine-vision algorithms

SEAN McCABE *(Koch)*



CARL WIENS (*illustration*)

can break down a picture into its constituent parts to identify faces and other objects. Speech-recognition algorithms parse and understand natural human speech. Speech synthesizers take the endless strings of 0's and 1's in which computers communicate and turn them into meaningful speech.

The results are becoming visible outside of computer science laboratories. Machines bested humans in chess years ago, and human leadership in *Jeopardy* was lost in 2011. Computers translate text from one language into dozens of others and have driven cars over more than 300,000 miles of open roads.

The pace at which machine performance has improved in these early days of the third millennium is staggering—so much so that some pundits predict the imminent advent of true artificial intelligence (AI). It may even be possible to contemplate the arrival in the not too distant future of a digital simulacrum of human-level intelligence. Computers may be getting nearer to receiving a passing grade on the so-called Turing test, conceived by British logician and patriarch of computing Alan Turing in 1950 as a means to discern whether a machine can really think. If a human judge cannot tell whether an answer to a question on any topic came from a computer program or a concealed human, the entity supplying the response must be considered intelligent.

In 1990 Hugh Loebner started an annual competition with a prize of $25,000 going to the first program that fools the (human) judges. So far no team has collected the Loebner Prize. The transcripts of these competitions make for hilarious reading, as human foils try to trip up judges by giving outrageous answers reminiscent of dialogue at a cocktail party for mushroom-ingesting dadaists.

Most academic and industry experts agree that an AI comparable to the intelligence of a typical adult—technology that can learn, infer and generalize in the way we do every day—remains a distant dream. Present-day software can't deal with complex linguistic utterances. It can't figure out that Noam Chomsky's paradigmatic "Colorless green ideas sleep furiously" is meaningless or that James Joyce's "The heaventree of stars hung with humid nightblue fruit" is an eloquent phrase that works its magic by the richness of its imagery. Of course, many people would find these sentences challenging unless they forgo any attempt to apply logic and just delight in the sensuousness of the word pictures, a capability far beyond the reach of any computer.

Setting aside these caveats, machine learning is the hottest technique in the market driving big data analytics. Its practitioners are in high demand, and the technology has been embraced enthusiastically by universities, defense and intelligence agencies, and companies—not just obvious ones like Google, Facebook and Amazon but also Walmart, Target and hedge funds. It remains to be seen whether refining the current crop of machine-learning algorithms will be sufficient for human-level speech or whether fundamental, Nobel Prize–winning breakthroughs will be essential. What is certain, though, is that unlike other standard-fare sci-fi predictions—faster than light warp drive, time travel or radical life-span extension—a Samantha-like verbal intelligence will be born within the lifetimes of many readers of this column.

But Would They Be Conscious?

Just because an AI program can talk like a smart and hyperefficient woman with a seductive voice does not imply that the program feels anything or is conscious. That's not to say that people wouldn't react, as Theodore did, by behaving as if the program had actual feelings for them. We have an innate tendency to impute feelings to many things, from our canine and feline companions to teddy bears, dolls, cars and other inanimate objects. That is the psychological reality of the human condition, which is why Samantha and her male-voiced counterpart would be a huge commercial

Lower Computing Costs Make AI Smarter

Here's how much the computational power of an iPhone 5s—18,200 MIPS (millions of instructions per second)—would have cost in inflation-adjusted dollars at the debut dates of milestone technologies, such as ENIAC.

As processing costs have dropped, the sophistication of artificial-intelligence has risen.

COMPUTING COSTS EQUAL TO AN IPHONE 5S

$45.14 trillion ENIAC (*1946*)

$26.37 trillion UNIVAC I (*1951*)

$1.34 trillion IBM 7090 (*1959*)

AI MILESTONES

1948: Norbert Wiener coins the term "cybernetics"

1958: John McCarthy invents the Lisp programming language at M.I.T.

1965: Joseph Weizenbaum invents the Eliza program to carry on a conversation, with a computer, which achieves fame as a means of dispensing psychotherapy.

1969: Shakey the robot demonstrates locomotion, perception and problem solving

success if they eventually make their way into the marketplace.

But that does not detract from the ontological question: Is simulating the relevant behavior—attraction, passion, desire, betrayal, angst, and so on—the same as having these feelings? The traditional answer is no. God endowed us humans, and only us, with an immortal living soul—without a soul, there is no consciousness.

Of course, we children of the Enlightenment know better. Consciousness is a product of the most highly organized chunk of matter in the universe, the central nervous system. And once the brain ceases to function, the conscious mind likewise dissolves. To put it as succinctly as a Zen koan: no brain, never mind.

On the bright side, this contemporary view also implies that if all the relevant neural mechanisms that underlie consciousness were to be faithfully replicated in an artificial brain, then this construct would be conscious. Function follows from mechanism as long as all the interactions involved in biological cause and effect are present.

Consider Scarlett Johansson, the actress who voiced Samantha, and a yet to be invented technique that could somehow scan her brain without harming it and map its 100 billion neurons and quadrillion synapses. From this scan, future neuroengineers might construct a gigantic software package, SimSamantha. It would need to run on a supercomputer that mathematically simulated the biochemical and biophysical activity of Johansson's brain.

If the neuroscientists had accurately captured all of her brain's bioelectric activity, SimSamantha would replicate the behavior of Johansson. But that's not all. According to the commonly held view, if this simulation were to capture all aspects of brain processing relevant to consciousness, then the computer program would experience the infatuation and rapture of being entranced by somebody else. SimSamantha would know real love.

In this line of argument, known as functionalism, the conscious mind is nothing but the brain at work—brain circuits pulsing on and off give rise to perception and thought. So functionalists, just like Theodore, believe that SimSamantha is capable of feeling, of loving him. (Whether she could simultaneously love 641 other people, as she claims, is a thornier matter.)

Functionalism is part of the miasma that hovers over and sometimes obscures the thinking of computer scientists and software engineers in Silicon Valley. The belief that sooner or later computers will become conscious is widespread among the digerati and engineers. But in thinking about one of the most difficult problems in all of science, brain researchers have come up with theories of consciousness that break with the functionalist tradition. The integrated information theory (IIT) of psychiatrist and neuroscientist Giulio Tononi offers a strikingly different perspective. (I previously described IIT in "Ubiquitous Minds," Scientific American Mind, January/February 2014.)

I am partial to IIT as the most plausible theory of consciousness and have worked with Tononi on aspects of it. The theory postulates that conscious experience arises from what Tononi terms "integrated information"—the multitude of sensory, motor and cognitive processes that are tied together to form the basis of any one subjective experience.

Any system that possesses some integrated information experiences something. This emphasis on integration reflects a fundamental characteristic of conscious experiences. Each one is highly integrated, holistic. As you are watching the colorful minimalist furniture and futuristic architecture in *Her*, you can't suddenly force yourself to see it in black-and-white. Its color is an integrated part of your experience. Whatever information you are conscious of is presented to you wholly and completely; it cannot be

SOURCE: HANS MORAVEC *The Robotics Institute, Carnegie Mellon University (cost of computing data)*; TIME & LIFE PICTURES *Getty Images (Shakey)*

$4.3 billion
ALTAIR 8800 (*1975*)

$130.7 million
COMMODORE 64 (*1982*)

$58.2 million
MAC II (*1987*)

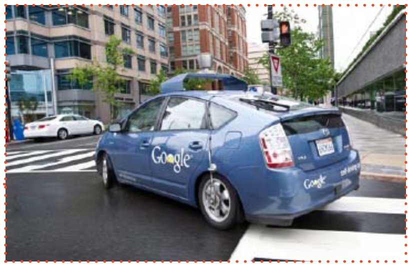
$222,560
GATEWAY G6-200 (*1997*)

$600
iPhone 5s (*2013*)

1985: Rodney Brooks introduces behavior-based robotics

1997: Deep Blue defeats Garry Kasparov

2009: Google builds a self-driving car



2011: IBM Watson defeats human *Jeopardy* champions

2011: Apple's Siri debuts

KAREN BLEIER *Getty Images* (*Google car*); SETH WENIG *Corbis* (*Jeopardy*)

subdivided. Underlying this unity of consciousness is a multitude of cause-and-effect interactions among the relevant parts of your brain.

Phi Meters

Integrated information can be calculated by considering your brain in a particular state. Taking into account the brain's immediate past and future, the theory computes a number that indicates how irreducible the brain is, that is, how much it resists being broken down into component parts. The bigger this number, denoted by the Greek letter Φ, or phi (pronounced "fi"), the broader and more sophisticated the conscious experience of the brain. If the organism has many neurons that are amply endowed with synaptic connections, Φ will be high. If the system is reducible to smaller, independent, noninteracting parts, Φ is zero. It has no experience at all. *Nada, rien,* nothing.

The brain of a patient in whom the entire corpus callosum—the 200 million fibers that connect the left cerebral hemisphere with the right—has been surgically cut to prevent epileptic seizures from spreading can be reduced to two independent hemispheres, each of which is conscious by itself. The once whole brain of the split-brain patient is now reduced to zero Φ because the shared contents of the two halves have been sundered. Meanwhile the two hemispheres are now endowed with nonzero Φ, a measure of consciousness present in each half.

To capture the experience of Samantha, it is essential to replicate the entire repertoire of interactions within her brain—what philosophers call its intrinsic causal properties—not just its input-output behavior, such as hearing and speech. This task can be accomplished only by building a faithful copy of her real brain using wires, transistors and other devices that have exactly the same cause-and-effect relationships among all components as in the real brain. A hypothetical artificial organ—call it BrainSamantha—that reflected the physical interactions among neurons, with one nerve cell changing the way another functions, would reproduce the same experiences as Samantha's brain.

The situation for SimSamantha software running on a digital computer would be quite different, however. The intrinsic causal properties of this program—how any circuit element in the computer switches on or off—are not the same as those of the biological brain being imitated. Ultimately what the computer does is shuffle binary charges from one transistor to a handful of others rather than sending electrical activity from one neuron to thousands of others. Paradoxically, SimSamantha would have the same ability to hear and speak as BrainSamantha, but without any feelings at all. A simulation is not the same as building an exact model. IIT stipulates that consciousness is an inherent feature of a highly complex set of interactions, taking into account the changes occurring within the system itself, not just the output of its processing. Consciousness cannot be reduced to something more elemental.

The causal properties of digital simulations are very different from BrainSamantha's. A computer program that simulates the weather illustrates what is missing. Although it can accurately forecast an approaching rainstorm, it will never be soaked with rain inside the computer (fortunately). And so it is with consciousness. While Theodore could not tell SimSamantha from BrainSamantha, only the latter could truly love him. Only the latter is endowed with consciousness, with true human feelings. **M**

FURTHER READING

- **From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0.** Masafumi Oizumi, Larissa Albantakis and Giulio Tononi in *PLOS Computational Biology* (in press).