Mind vs. Machine

IN THE RACE TO BUILD COMPUTERS THAT CAN THINK LIKE HUMANS, THE PROVING GROUND IS THE TURING TEST —AN ANNUAL BATTLE BETWEEN THE WORLD'S MOST ADVANCED ARTIFICIAL-INTELLIGENCE PROGRAMS AND ORDINARY PEOPLE. THE OBJECTIVE? TO FIND OUT WHETHER A COMPUTER CAN ACT "MORE HUMAN" THAN A PERSON. IN HIS OWN QUEST TO BEAT THE MACHINES, THE AUTHOR DISCOVERS THAT THE MARCH OF TECHNOLOGY ISN'T JUST CHANGING HOW WE LIVE, IT'S RAISING NEW QUESTIONS ABOUT WHAT IT MEANS TO BE HUMAN.

By Brian Christian

BRIGHTON, ENGLAND, SEPTEMBER 2009. I wake up in a hotel room 5,000 miles from my home in Seattle. After breakfast, I step out into the salty air and walk the coastline of the country that invented my language, though I find I can't understand a good portion of the signs I pass on my way—LET AGREED, one says, prominently, in large print, and it means nothing to me.

I pause, and stare dumbly at the sea for a moment, parsing and reparsing the sign. Normally these kinds of linguistic curiosities and cultural gaps intrigue me; today, though, they are mostly a cause for concern. In two hours, I will sit down at a computer and have a series of five-minute instant-message chats with several strangers. At the other end of these chats will be a psychologist, a linguist, a computer scientist, and the host of a popular British technology show. Together they form a judging panel, evaluating my ability to do one of the strangest things I've ever been asked to do.

I must convince them that I'm human.

Fortunately, I *am* human; unfortunately, it's not clear how much that will help.

Also see:

From Luddites to Predators, Men vs. Machines Through Time
Humanity's fears and dilemmas resulting from technology since the Industrial Revolution.

Technology and Humanity in *The Atlantic*
Writings on the interface between technology and humanity by Mark Twain, Oliver Wendell Holmes,

Nobel Laureate James D. Watson, James Fallows, and others.

The Turing Test

Each year for the past two decades, the artificial-intelligence community has convened for the field's most anticipated and controversial event—a meeting to confer the Loebner Prize on the winner of a competition called the Turing Test. The test is named for the British mathematician Alan Turing, one of the founders of computer science, who in 1950 attempted to answer one of the field's earliest questions: can machines think? That is, would it ever be possible to construct a computer so sophisticated that it could actually be said to be thinking, to be intelligent, to have a mind? And if indeed there were, someday, such a machine: how would we know?

Instead of debating this question on purely theoretical grounds, Turing proposed an experiment. Several judges each pose questions, via computer terminal, to several pairs of unseen correspondents, one a human "confederate," the other a computer program, and attempt to discern which is which. The dialogue can range from small talk to trivia questions, from celebrity gossip to heavy-duty philosophy—the whole gamut of human conversation. Turing predicted that by the year 2000, computers would be able to fool 30 percent of human judges after five minutes of conversation, and that as a result, one would "be able to speak of machines thinking without expecting to be contradicted."

Turing's prediction has not come to pass; however, at the 2008 contest, the top-scoring computer program missed that mark by just a single vote. When I read the news, I realized instantly that the 2009 test in Brighton could be the decisive one. I'd never attended the event, but I felt I had to go—and not just as a spectator, but as part of the human defense. A steely voice had risen up inside me, seemingly out of nowhere: *Not on my watch*. I determined to become a confederate.

The thought of going head-to-head (head-to-motherboard?) against some of the world's top AI programs filled me with a romantic notion that, as a confederate, I would be *defending the human race*, à la Garry Kasparov's chess match against Deep Blue.

During the competition, each of four judges will type a conversation with one of us for five minutes, then the other, and then will have 10 minutes to reflect and decide which one is the human. Judges will also rank all the contestants—this is used in part as a tiebreaking measure. The computer program receiving the most votes and highest ranking from the judges (regardless of whether it passes the Turing Test by fooling 30 percent of them) is awarded the title of the Most Human Computer. It is this title that the research teams are all gunning for, the one with the cash prize (usually $3,000), the one with which most everyone involved in the contest is principally concerned. But there is also, intriguingly, another title, one given to the *confederate* who is most convincing: the Most Human Human award.

One of the first winners, in 1994, was the journalist and science-fiction writer Charles Platt. How'd he do it? By "being moody, irritable, and obnoxious," as he explained in *Wired* magazine—which strikes me as not only hilarious and bleak, but, in some deeper sense, a call to arms: how, in fact, do we be the

most human we can be—not only under the constraints of the test, but in life?

THE IMPORTANCE OF BEING YOURSELF

Since 1991, the Turing Test has been administered at the so-called Loebner Prize competition, an event sponsored by a colorful figure: the former baron of plastic roll-up portable disco dance floors, Hugh Loebner. When asked his motives for orchestrating this annual Turing Test, Loebner cites laziness, of all things: his utopian future, apparently, is one in which unemployment rates are nearly 100 percent and virtually all of human endeavor and industry is outsourced to intelligent machines.

To learn how to become a confederate, I sought out Loebner himself, who put me in touch with contest organizers, to whom I explained that I'm a nonfiction writer of science and philosophy, fascinated by the Most Human Human award. Soon I was on the confederate roster. I was briefed on the logistics of the competition, but not much else. "There's not much more you need to know, really," I was told. "You are human, so just be yourself."

Just be yourself has become, in effect, the confederate motto, but it seems to me like a somewhat naive overconfidence in human instincts—or at worst, like fixing the fight. Many of the AI programs we confederates go up against are the result of decades of work. Then again, so are we. But the AI research teams have huge databases of test runs for their programs, and they've done statistical analysis on these archives: the programs know how to deftly guide the conversation away from their shortcomings and toward their strengths, know which conversational routes lead to deep exchange and which ones fizzle. The average off-the-street confederate's instincts—or judge's, for that matter—aren't likely to be so good. This is a strange and deeply interesting point, amply proved by the perennial demand in our society for dating coaches and public-speaking classes. The transcripts from the 2008 contest show the humans to be such wet blankets that the judges become downright apologetic for failing to provoke better conversation: "I feel sorry for the humans behind the screen, I reckon they must be getting a bit bored talking about the weather," one writes; another offers, meekly, "Sorry for being so banal." Meanwhile a computer appears to be charming the pants off one judge, who in no time at all is gushing *LOL*s and smiley-face emoticons. We can do better.

Thus, my intention from the start was to thoroughly disobey the advice to just show up and be myself—I would spend months preparing to give it everything I had.

Ordinarily this notion wouldn't be odd at all, of course—we train and prepare for tennis competitions, spelling bees, standardized tests, and the like. But given that the Turing Test is meant to evaluate *how human* I am, the implication seems to be that being human (and being oneself) is about more than simply showing up.

THE SENTENCE

To understand why our human sense of self is so bound up with the history of computers, it's important to realize that computers used to *be human*. In the early 20th century, before a "computer"

was one of the digital processing devices that permeate our 21st-century lives, it was something else: a job description.

From the mid-18th century onward, computers, many of them women, were on the payrolls of corporations, engineering firms, and universities, performing calculations and numerical analysis, sometimes with the use of a rudimentary calculator. These original, human computers were behind the calculations for everything from the first accurate prediction, in 1757, for the return of Halley's Comet—early proof of Newton's theory of gravity—to the Manhattan Project at Los Alamos, where the physicist Richard Feynman oversaw a group of human computers.

It's amazing to look back at some of the earliest papers on computer science and see the authors attempting to explain what exactly these new contraptions were. Turing's paper, for instance, describes the unheard-of "digital computer" by making analogies to a *human* computer:

> The idea behind digital computers may be explained by saying that these machines are intended to carry out any operations which could be done by a human computer.

Of course, in the decades that followed, we know that the quotation marks migrated, and now it is "digital computer" that is not only the default term, but the *literal* one. In the mid-20th century, a piece of cutting-edge mathematical gadgetry was said to be "like a computer." In the 21st century, it is the human math whiz who is "like a computer." It's an odd twist: we're *like* the thing that used to be *like* us. We imitate our old imitators, in one of the strange reversals in the long saga of human uniqueness.

Philosophers, psychologists, and scientists have been puzzling over the essential definition of human uniqueness since the beginning of recorded history. The Harvard psychologist Daniel Gilbert says that every psychologist must, at some point in his or her career, write a version of what he calls "The Sentence." Specifically, The Sentence reads like this:

> The human being is the only animal that _____.

The story of humans' sense of self is, you might say, the story of failed, debunked versions of The Sentence. Except now it's not just the animals that we're worried about.

We once thought humans were unique for using language, but this seems less certain each year; we once thought humans were unique for using tools, but this claim also erodes with ongoing animal-behavior research; we once thought humans were unique for being able to do mathematics, and now we can barely imagine being able to do what our calculators can.

We might ask ourselves: Is it appropriate to allow our definition of our own uniqueness to be, in some sense, *reactive* to the advancing front of technology? And why is it that we are so compelled to feel unique in the first place?

"Sometimes it seems," says Douglas Hofstadter, a Pulitzer Prize–winning cognitive scientist, "as though each new step towards AI, rather than producing something which everyone agrees is real intelligence, merely reveals what real intelligence is *not*." While at first this seems a consoling position—one that keeps our unique claim to thought intact—it does bear the uncomfortable appearance of a gradual retreat, like a medieval army withdrawing from the castle to the keep. But the retreat can't continue indefinitely. Consider: if everything that we thought hinged on thinking turns out to not involve it, then … what is thinking? It would seem to reduce to either an epiphenomenon—a kind of "exhaust" thrown off by the brain—or, worse, an illusion.

Where is the keep of our *selfhood*?

The story of the 21st century will be, in part, the story of the drawing and redrawing of these battle lines, the story of *Homo sapiens* trying to stake a claim on shifting ground, flanked by beast and machine, pinned between meat and math.

Is this retreat a good thing or a bad thing? For instance, does the fact that computers are so good at mathematics in some sense *take away* an arena of human activity, or does it *free* us from having to do a nonhuman activity, liberating us into a more human life? The latter view seems to be more appealing, but less so when we begin to imagine a point in the future when the number of "human activities" left for us to be "liberated" into has grown uncomfortably small. What then?

Alan Turing proposed his test as a way to measure technology's progress, but it just as easily lets us measure our own. The Oxford philosopher John Lucas says, for instance, that if we fail to prevent the machines from passing the Turing Test, it will be "not because machines are so intelligent, but because humans, many of them at least, are so wooden."

Beyond its use as a technological benchmark, the Turing Test is, at bottom, about the act of communication. I see its deepest questions as practical ones: How do we connect meaningfully with each other, as meaningfully as possible, within the limits of language and time? How does empathy work? What is the process by which someone enters into our life and comes to mean something to us? These, to me, are the test's most central questions—the most central questions of being human.

Part of what's fascinating about studying the programs that have done well at the Turing Test is seeing how conversation can work in the total absence of emotional intimacy. A look at the transcripts of Turing Tests past is, frankly, a sobering tour of the various ways in which we demur, dodge the question, lighten the mood, change the subject, distract, burn time: what shouldn't pass for real conversation at the Turing Test probably shouldn't be allowed to pass for real conversation in everyday life either.

HOW TO TALK LIKE A HUMAN

When the Turing Test was first proposed in 1950, it was a hypothetical: technology was nowhere near the point at which a practical test was possible. But, as we know, it got there; the first conversational

computer program to attract significant notice and attention was Eliza, written in 1964 and 1965 by Joseph Weizenbaum at MIT. Modeled after a Rogerian therapist, Eliza worked on a very simple principle: extract key words from the users' own language, and pose their statements back to them. ("I am unhappy." "Do you think coming here will help you not to be unhappy?") If in doubt, it might fall back on some completely generic phrases, like "Please go on." This technique of fitting the users' statements into predefined patterns and responding with a prescribed phrasing of its own—called "template matching"—was Eliza's only capacity.

A look at an Eliza transcript reveals how adeptly such an impoverished set of rules can, in the right context, pass at a glance for understanding:

User: Men are all alike.

Eliza: In what way?

User: They're always bugging us about something or other.

Eliza: Can you think of a specific example?

User: Well, my boyfriend made me come here.

Eliza: Your boyfriend made you come here.

User: He says I'm depressed much of the time.

Eliza: I am sorry to hear you are depressed.

User: It's true. I am unhappy.

Eliza: Do you think coming here will help you not to be unhappy?

User: I need some help, that much seems certain.

Eliza: What would it mean to you if you got some help?

User: Perhaps I could learn to get along with my mother.

Eliza: Tell me more about your family.

The results were stunning: many of the people who first talked with Eliza were convinced that they were having a genuine human exchange. In some cases, even Weizenbaum's own insistence to the contrary was of no use. People asked to be left alone to talk "in private," sometimes for hours, and returned with reports of having had a meaningful therapeutic experience. Meanwhile, academics leapt to conclude that Eliza represented "a general solution to the problem of computer understanding of natural language."

One of the strangest twists to the Eliza story, however, was the reaction of the *medical* community, which decided Weizenbaum had hit upon something both brilliant and useful. *The Journal of Nervous*

and Mental Disease, for example, said of Eliza in 1966:

> Several hundred patients an hour could be handled by a computer system designed for this purpose. The human therapist, involved in the design and operation of this system, would not be replaced, but would become a much more efficient man.

The famed scientist Carl Sagan, in 1975, concurred:

> I can imagine the development of a network of computer psychotherapeutic terminals, something like arrays of large telephone booths, in which, for a few dollars a session, we would be able to talk with an attentive, tested, and largely non-directive psychotherapist.

As for Weizenbaum, appalled and horrified, he did something almost unheard-of: an about-face on his entire career. He pulled the plug on the Eliza project, encouraged his own critics, and became one of science's most outspoken opponents of AI research. But the genie was out of the bottle, and there was no going back. The basic "template matching" skeleton and approach of Eliza has been reworked and implemented in some form or another in almost every chat program since, including the contenders at the 2009 Loebner Prize competition. The enthusiasm—as well as the unease—about these programs has only grown.

THE JERK AS GENIUS

The first Loebner Prize competition was held on November 8, 1991, at the Boston Computer Museum. In its first few years, the contest required each program and human confederate to choose a topic, as a means of limiting the conversation. One of the confederates in 1991 was the Shakespeare expert Cynthia Clay, who was, famously, deemed a computer by three different judges after a conversation about the playwright. The consensus seemed to be: "No one knows that much about Shakespeare." (For this reason, Clay took her misclassifications as a compliment.)

The programmer Joseph Weintraub chose "Whimsical Conversation" for his PC Therapist III and went on to earn the very first Most Human Computer award. As the program practiced it, "whimsical conversation" was not a topic at all, but a manner—a crucial distinction. Weintraub was able to play off a looseness in the contest rules, the natural behavior patterns of chatbots, and people's stereotypes about computers and humans, all at once. Strategically, this was brilliant.

By dutifully and scrupulously providing information in response to the questions asked, Clay demonstrated her knowledge and understanding—but sometimes in a factual, encyclopedic way commonly associated with computer systems. Here is a sample of Clay's conversation:

Judge: What is your opinion on Shakespeare's plays?

Confederate: That's pretty general; would you be more specific? Otherwise, I'll just say I like them.

The computer at the "Whimsical Conversation" terminal stood in stark contrast:

Judge: It has been a long day! Now I think I'm ready for some fun. The weather isn't very pleasant today. Do you have any jokes that could lift the clouds?

Computer: Everybody talks about the weather but nobody seems to do much about it. Why do you need to tell me now you think you're ready for some fun? Groannnnnn … … … … …

Judge: What is the definition of whimsical conversation?

Computer: Didn't you say well anyway I would imagine whimsical conversation is like a series of knight's moves in chess—skrt of zig-zag rather than following any nparticular line.? [Weintraub's program, like many that followed it, faked typos. Any unintentional typing errors in the transcripts that follow have been corrected.] Why do you ask? Was that a clever comment, or what? Are such questions much on your mind?

Judge: I like the image of knights moving haphazardly across the chess board, does that mean there is no thought to whimsical conversation?

Computer: I could swear you just said how do you define whimsical?? How do yku define whimsical? Were you always so sick sick sick? Why do you need to tell me you like the image of knights moving haphazardly across the chess board?

"I really thought [PC Therapist] was human, because it … was weird and funny, in a normal sort of way," said one of the judges, a Harvard graduate student. Weintraub's program, shifting topics wildly and spouting non sequiturs and canned one-liners, came off as zany, a jokester, a much more "human" personality type. At least I used to think so—before I learned how easy this was to mimic.

As Richard Wallace, three-time winner of the Most Human Computer award ('00, '01, and '04), explains:

> Experience with [Wallace's chatbot] ALICE indicates that most casual conversation is "stateless," that is, each reply depends only on the current query, without any knowledge of the history of the conversation required to formulate the reply.

Many human conversations function in this way, and it behooves AI researchers to determine which types of conversation are stateless—with each remark depending only on the last—and try to create these very sorts of interactions. It's our job as confederates, as humans, to resist them.

One of the classic stateless conversation types is the kind of zany free-associative riffing that Weintraub's program, PC Therapist III, employed. Another, it turns out, is verbal abuse.

In May 1989, Mark Humphrys, a 21-year-old University College Dublin undergraduate, put online an Eliza-style program he'd written, called "MGonz," and left the building for the day. A user (screen name "Someone") at Drake University in Iowa tentatively sent the message "finger" to Humphrys's account—an early-Internet command that acted as a request for basic information about a user. To

Someone's surprise, a response came back immediately: "cut this cryptic shit speak in full sentences." This began an argument between Someone and MGonz that lasted almost an hour and a half. (The best part was undoubtedly when Someone said, "you sound like a goddamn robot that repeats everything.")

Returning to the lab the next morning, Humphrys was stunned to find the log, and felt a strange, ambivalent emotion. His program might have just shown how to pass the Turing Test, he thought—but the evidence was so profane that he was afraid to publish it.

Humphrys's twist on the Eliza paradigm was to abandon the therapist persona for that of an abusive jerk; when it lacked any clear cue for what to say, MGonz fell back not on therapy clichés like "How does that make you feel?" but on things like "You are obviously an asshole," or "Ah type something interesting or shut up." It's a stroke of genius because, as becomes painfully clear from reading the MGonz transcripts, argument is stateless—that is, unanchored from all context, a kind of Markov chain of riposte, meta-riposte, meta-meta-riposte. Each remark after the first is only about the previous remark. If a program can induce us to sink to this level, of course it can pass the Turing Test.

Once again, the question of what types of human behavior computers can imitate shines light on how we conduct our own, human lives. Verbal abuse is simply less complex than other forms of conversation. In fact, since reading the papers on MGonz, and transcripts of its conversations, I find myself much more able to constructively manage heated conversations. Aware of the stateless, knee-jerk character of the terse remark I want to blurt out, I recognize that that remark has far more to do with a reflex reaction to the very last sentence of the conversation than with either the issue at hand or the person I'm talking to. All of a sudden, the absurdity and ridiculousness of this kind of escalation become quantitatively clear, and, contemptuously unwilling to act like a bot, I steer myself toward a more "stateful" response: better living through science.

BEWARE OF BANALITY

Entering the Brighton Centre, I found my way to the Loebner Prize contest room. I saw rows of seats, where a handful of audience members had already gathered, and up front, what could only be the bot programmers worked hurriedly, plugging in tangles of wires and making the last flurries of keystrokes. Before I could get too good a look at them, this year's test organizer, Philip Jackson, greeted me and led me behind a velvet curtain to the confederate area. Out of view of the audience and the judges, the four of us confederates sat around a rectangular table, each at a laptop set up for the test: Doug, a Canadian linguistics researcher; Dave, an American engineer working for Sandia National Laboratories; Olga, a speech-research graduate student from South Africa; and me. As we introduced ourselves, we could hear the judges and audience members slowly filing in, but couldn't see them around the curtain. A man zoomed by in a green floral shirt, talking a mile a minute and devouring finger sandwiches. Though I had never met him before, I knew instantly he could be only one person: Hugh Loebner. Everything was in place, he told us, between bites, and the first round of the test would start momentarily. We four confederates grew quiet, staring at the blinking cursors on our laptops. My

hands were poised over the keyboard, like a nervous gunfighter's over his holsters.

The cursor, blinking. I, unblinking. Then all at once, letters and words began to materialize:

Hi how are you doing?

The Turing Test had begun.

I had learned from reading past Loebner Prize transcripts that judges come in two types: the small-talkers and the interrogators. The latter go straight in with word problems, spatial-reasoning questions, deliberate misspellings. They lay down a verbal obstacle course, and you have to run it. This type of conversation is extraordinarily hard for programmers to prepare against, because anything goes—and this is why Turing had language and conversation in mind as his test, because they are really a test of everything. The downside to the give-'em-the-third-degree approach is that it doesn't leave much room to express yourself, personality-wise.

The small-talk approach has the advantage of making it easier to get a sense of who a person is—if you are indeed talking to a person. And this style of conversation comes more naturally to layperson judges. For one reason or another, small talk has been explicitly and implicitly encouraged among Loebner Prize judges. It's come to be known as the "strangers on a plane" paradigm. The downside is that these conversations are, in some sense, uniform—familiar in a way that allows a programmer to anticipate a number of the questions.

I started typing back.

Confederate: hey there!

Confederate: i'm good, excited to actually be typing

Confederate: how are you?

I could imagine the whole lackluster conversation spread out before me: *Good. Where are you from? / Seattle. How about yourself? / London.*

Four minutes and 43 seconds left. My fingers tapped and fluttered anxiously.

I could just feel the clock grinding away while we lingered over the pleasantries. I felt this desperate urge to go off script, cut the crap, cut to the chase—because I knew that the computers could do the small-talk thing, which played directly into their preparation. As the generic civilities stretched forebodingly out before me, I realized that this very kind of conversational boilerplate was the enemy, every bit as much as the bots. *How*, I was thinking as I typed another unassuming pleasantry, *do I get an obviously human connection to happen?*

Taking Turns

Part of what I needed to figure out was how to exploit the Loebner Prize's unusual "live typing"

medium. The protocol being used was unlike e-mails, text messages, and standard instant-messaging systems in a very crucial way: it transmitted our typing keystroke by keystroke. The judge and I were watching each other type, typos and backspacing and all. I remember some Internet chat programs back in the '90s trying out this character-at-a-time approach, but people for the most part rejected it. It was too invasive, was the feeling: what people like about writing is the time and space to compose and edit a message before sharing it with the other person. The advantage of the character-at-a-time transmission, though, is that it approaches much more closely the condition of speech, with its fluidity of turn-taking and its choppy grammar: what's lost in eloquence is made up for in agility.

It also, then, lets us see typing's "negative space": hesitation. In a chat conversation where text is transmitted with every carriage return, only egregiously long pauses are taken to be part of the interaction. With more fluid and immediate feedback, silence acquires meaning. Failing to quickly answer a question in a face-to-face conversation, for instance, is tantamount in many cases to answering it. I recall asking a friend of mine how things were going with the woman he'd started seeing; the *um* and the tiny pause in which he searched for the right words made it clear they were having problems.

So what did the Loebner Prize's unusual (and recently implemented) protocols enable and disable, compared with the standard, walkie-talkie, turn-taking style? Long missives weren't going to work, as they had in previous years, when programs were able to steamroll the judges by eating up the clock and delivering ridiculously prolix answers. If two parties are taking strict turns under time pressure, they are putting themselves at the mercy of the length of the other's turns. Here, for instance, is one program's conversation transcript from 1996:

Computer: How about telling me a joke?

Judge: No.

Computer: Here's my favorite joke ... A dog limps into a bar with a bloody bandage on his foot. He's wearing a 10-gallon hat, six-shooters in his holsters and chaps. He growled, "I'm a-lookin' for the man that shot my paw."

"The joke's not funny ..." the judge writes, giving the program an opening to tell another one—which it does ("A knotty, worn-out old string walks into a bar ..."). Meanwhile, three full minutes have elapsed. If a computer (or confederate) started rambling on too long under the new, live-typing protocols, the judge could and would just cut it off.

And so another piece of my confederate strategy fell into place. I would treat the Turing Test's strange and unfamiliar textual medium more like spoken English, and less like the written language. I would attempt to disrupt the turn-taking "wait and parse" pattern that computers understand, and create a single, flowing duet of verbal behavior, emphasizing timing. If computers understand little about verbal "harmony," they understand even less about rhythm.

If nothing was happening on my screen, whether or not it was my turn, I'd elaborate a little on my answer, or add a parenthetical, or throw a question back at the judge—just as we offer and/or fill audible silence when we talk out loud. If the judge took too long considering the next question, I'd keep talking. I would be the one (unlike the bots) with something to prove. If I knew what the judge was about to write, I'd spare him the keystrokes and jump in.

There's a trade-off, of course, between the number of opportunities for serve and volley, and the sophistication of the responses themselves. The former thrives with brevity, the latter with length. It seemed to me, though, that so much of the nuance (or difficulty) in conversation comes from understanding (or misunderstanding) a question and offering an appropriate (or inappropriate) response—thus, it made sense to maximize the number of interchanges.

Some judges, I discovered, would be startled or confused at this jumping of the gun, and I saw them pause, hesitate, yield, even start backspacing what they had half-written. Other judges cottoned on immediately, and leapt right in after me.

In the first round of the 2009 contest, judge Shalom Lappin—a computational linguist at King's College London—spoke with a computer program called Cleverbot, and then with me. My strategy of verbosity was clearly in evidence: I made 1,089 keystrokes in five minutes (3.6 keystrokes a second) to Cleverbot's 356 (1.2/sec), and Lappin made 548 keystrokes (1.8/sec) in my conversation, compared with 397 (1.3/sec) with Cleverbot. Not only did I say three times as much as my silicon adversary, but I engaged the judge more, to the tune of 38 percent more typing from Lappin.

Reviewing the logs later, though, I looked for a way to quantify the fluidity of the human-to-human interactions against the rigidity of the human-to-machine ones. It occurred to me that you could count the changes—let's call them "swaps"—of the party who typed the latest keystroke. I did manage to type three times as much as Cleverbot, but the real story, as it turns out, is in the swaps. Lappin's conversation with Cleverbot had 33; his conversation with me had 492, almost 15 times as many.

The Confederate's gambit

You don't *converse with* Google, or with most computer systems; you *depose* them. There's a crucial difference. Interestingly, many Loebner Prize judges approach the Turing Test as a kind of cross-examination; strangely, a number of confederates also seem to approach it with that dynamic in mind. One of the human-to-human conversations in 2008 never manages to get out of that stiff question-and-response mode:

Judge: Do you have a long drive?

Confederate: fairly long

Judge: so do I :(ah well, do you think you could have used public transport?

Confederate: i could have

Judge: and why not?

Confederate: i chose not to

Judge: that's fair. Do you think we have too many cars on the road or not enough today?

Confederate: it's not for me to say

Yawn! But the computer in this pair is playful with the judge from the get-go:

Judge: HI

Computer: Amen to that.

Judge: quite the evangelist

Computer: Our Father, who art in cyberspace, give us today our daily bandwidth.

Judge: evangelist / nerd lol. So how are things with you today?

And with that, the program has practically sealed up the judge's confidence in its humanity with its second sentence. Note that the confederate's stiff answers prompt more grilling and forced conversation—*what's your opinion on such-and-such political topic?* But with the computer, the judge, misled by the opening wisecracks into assuming it's the real person, remains utterly casual: *how are things?* This makes the contest easier for the computer and harder for the confederate.

It surprised me to see some confederates being coy with their judges. Asked what kind of engineer he is, Dave, to my left, answered, "A good one. :)" And Doug, to my right, responded to a question about what brought him to Brighton with "if I tell you, you'll know immediately that I'm human ;-)" For my money, wit is very successful, but coyness is a double-edged sword. You show a sense of humor, but you jam the cogs of the conversation. Probably the most dangerous thing a confederate can do in a Turing Test is stall. It's suspect—as the guilty party would tend to be the one running out the clock—and it squanders your most precious resource: time.

The humans in a Turing Test are strangers, limited to a medium that is slow and has no vocal tonality, and without much time. A five-second Turing Test would be an easy win for the machines: the judges, barely able to even say "hello," simply wouldn't be able to get enough data from their respondents to make any kind of judgment. A five-hour test would be an easy win for the humans. The Loebner Prize organizers have tried different time limits since the contest's inception, but in recent years they've mostly adhered to Turing's original prescription of five minutes: around the point when conversation starts to get interesting.

A big part of what I needed to do as a confederate was simply to make as much engagement happen in those minutes as I physically and mentally could. Rather than adopt the terseness of a deponent, I offered the prolixity of a writer. In other words, I talked *a lot*. I stopped typing only when to keep going would have seemed blatantly impolite or blatantly suspicious. The rest of the time, my fingers were

moving. I went out of my way to embody that maxim of "A bore is a man who, being asked 'How are you?' starts telling you how he is."

Judge: Hi, how's things?

Confederate: hey there

Confederate: things are good

Confederate: a lot of waiting, but ...

Confederate: good to be back now and going along

Confederate: how are you?

When we'd finished, and my judge was engaged in conversation with one of my computer counterparts, I strolled around the table, seeing what my comrades were up to. Looking over at my fellow confederate Dave's screen, I noticed his conversation began like he was on the receiving end of an interrogation, and he was answering in a kind of minimal staccato:

Judge: Are you from Brighton?

Confederate: No, from the US

Judge: What are you doing in Brighton?

Confederate: On business

Judge: How did you get involved with the competition?

Confederate: I answered an e-mail.

Like a good deponent, he let the questioner do all the work. When I saw how stiff Dave was being, I confess I felt a certain confidence—I, in my role as the world's worst deponent, was perhaps in fairly good shape as far as the Most Human Human award was concerned.

This confidence lasted approximately 60 seconds, or enough time for me to continue around the table and see what another fellow confederate, Doug, and his judge had been saying.

Judge: Hey Bro, I'm from TO.

Confederate: cool

Confederate: leafs suck

Confederate: ;-)

Judge: I am just back from a sabbatical in the CS Dept. at U of T.

Confederate: nice!

Judge: I remember when they were a great team.

Judge: That carbon date me, eh?

Confederate: well, the habs were a great team once, too ...

Confederate: *sigh*

Judge: YEH, THEY SUCK TOO.

Confederate: (I'm from Montreal, if you didn't guess)

Doug and his judge had just discovered that they were both Canadian. They let rip with abbreviations and nicknames and slang and local references. And then they started to talk about hockey.

I was in trouble.

SIX MONTHS AFTER the 2009 contest, a video appeared on YouTube of a man having a shockingly cogent conversation with a bot about Shakespeare's *Hamlet*. Some suspected it might herald a new age for chatbots, and for AI. Others, including myself, were unimpressed. Sophisticated behavior doesn't necessarily indicate a *mind*. It might indicate just a *memory*. As Dalí so famously put it, "The first man to compare the cheeks of a young woman to a rose was obviously a poet; the first to repeat it was possibly an idiot."

For instance, Richard Wallace, the three-time Most Human Computer winner, recounts an "AI urban legend" in which

> a famous natural language researcher was embarrassed ... when it became apparent to his audience of Texas bankers that the robot was consistently responding to the *next* question he was about to ask ... [His] demonstration of natural language understanding ... was in reality nothing but a simple script.

The moral of the story: no demonstration is ever sufficient. Only *interaction* will do. In the 1997 contest, one judge gets taken for a ride by Catherine, waxing political and really engaging in the topical conversation "she" has been programmed to lead about the Clintons and Whitewater. In fact, everything is going swimmingly until the very end, when the judge signs off:

Judge: it looks like i've got to go, Catherine. its been nice talking with you

Judge: do you need water or something before the 5 minute rounds?

At which point the program erupts into gibberish.

Computer: OK, yes on balance ... Time to get off this one I think and onto something more down to earth!

Judge: like what?

Computer: Almost certainly! let's move on

Computer: I suppose it depends on where you're coming from, but as the song goes in My Fair Lady, "Why can't a woman be more like a man?"

Defeat from the jaws of victory.

We so often think of intelligence, of AI, in terms of sophistication, or complexity of behavior. But in so many cases, it's impossible to say much with certainty about the program itself, because any number of different pieces of software—of wildly varying levels of "intelligence"—could have produced that behavior.

No, I think sophistication, complexity of behavior, is not it at all. For instance, you can't judge the intelligence of an orator by the eloquence of his prepared remarks; you must wait until the Q&A and see how he fields questions. The computation theorist Hava Siegelmann once described intelligence as "a kind of sensitivity to things." These Turing Test programs that hold forth may produce interesting output, but they're rigid and inflexible. They are, in other words, insensitive—occasionally fascinating talkers that cannot listen.

As computing technology in the 21st century moves increasingly toward mobile devices, we've seen the 1990s' explosive growth in processor speed taper off, and product development become less about raw computing horsepower than about the overall design of the product and its fluidity, reactivity, and ease of use. This fascinating shift in computing emphasis may be the cause, effect, or correlative of a healthier view of human intelligence—an understanding, not so much that it is complex and powerful, per se, as that it is reactive, responsive, sensitive, nimble. Our computers, flawed mirrors that they are, have helped us see that about ourselves.

The Most Human Human

The Most Human Computer award in 2009 goes to David Levy and his program, Do-Much-More. Levy, who also won in '97, with Catherine, is an intriguing guy: he was one of the big early figures in the digital-chess scene of the '70s and '80s, and was one of the organizers of the Marion Tinsley–Chinook checkers matches that preceded the Kasparov–Deep Blue showdowns in the '90s. He's also the author of the recent nonfiction book *Love and Sex With Robots*, to give you an idea of the sorts of things that are on his mind when he's not competing for the Loebner Prize.

Levy stands up, to applause, accepts the award from Philip Jackson and Hugh Loebner, and makes a short speech about the importance of AI for a bright future, and the importance of the Loebner Prize for AI. I know what's next on the agenda, and my stomach knots. I'm certain that Doug's gotten it; he and the judge were talking Canada 30 seconds into their conversation.

Ridiculous Canadians and their ice hockey, I'm thinking. Then I'm thinking how ridiculous it is that I'm even allowing myself to get this worked up about some silly award. Then I'm thinking how

ridiculous it is to fly 5,000 miles just to have a few minutes' worth of IM conversations. Then I'm thinking how maybe it'll be great to be the runner-up; I can compete again in 2010, in Los Angeles, with the home-field cultural advantage, and finally prove—

"And the results here show also the identification of the humans," Jackson announces, "and from the ranking list we can see that 'Confederate 1,' which is Brian Christian, was the most human."

And he hands me the certificate for the Most Human Human award.

I DIDN'T KNOW how to feel, exactly. It seemed strange to treat the award as meaningless or trivial, but did winning really represent something about me as a person? More than anything, I felt that together, my fellow confederates and I had avenged the mistakes of 2008 in dramatic fashion. That year, the 12 judges decided five times that computer programs were more human than confederates. In three of those instances, the judge was fooled by a program named Elbot, which was the handiwork of a company called Artificial Solutions, one of many new businesses leveraging chatbot technology. One more deception, and Elbot would have tricked 33 percent of that year's dozen judges—surpassing Turing's 30 percent mark, and making history. After Elbot's victory at the Loebner Prize and the publicity that followed, the company seemingly decided to prioritize the Elbot software's more commercial applications; at any rate, it had not entered the '09 contest as the returning champion.

In some ways a closer fight would have been more dramatic. Between us, we confederates hadn't permitted a single vote to go the machines' way. Whereas 2008 was a nail-biter, 2009 was a rout. We think of science as an unhaltable, indefatigable advance. But in the context of the Turing Test, humans —dynamic as ever—don't allow for that kind of narrative. We don't provide the kind of benchmark that sits still.

As for the prospects of AI, some people imagine the future of computing as a kind of heaven. Rallying behind an idea called "The Singularity," people like Ray Kurzweil (in *The Singularity Is Near*) and his cohort of believers envision a moment when we make smarter- than-us machines, which make machines smarter than themselves, and so on, and the whole thing accelerates exponentially toward a massive ultra-intelligence that we can barely fathom. Such a time will become, in their view, a kind of a techno-Rapture, in which humans can upload their consciousness onto the Internet and get assumed—if not bodily, than at least mentally—into an eternal, imperishable afterlife in the world of electricity.

Others imagine the future of computing as a kind of hell. Machines black out the sun, level our cities, seal us in hyperbaric chambers, and siphon our body heat forever.

I'm no futurist, but I suppose if anything, I prefer to think of the long-term future of AI as a kind of purgatory: a place where the flawed but good-hearted go to be purified—and tested—and come out better on the other side.

Who would have imagined that the computer's earliest achievements would be in the domain of logical analysis, a capacity once held to be what made us most different from everything else on the planet?

That it could fly a plane and guide a missile before it could ride a bike? That it could create plausible preludes in the style of Bach before it could make plausible small talk? That it could translate before it could paraphrase? That it could spin half-discernible essays on postmodern theory before it could be shown a chair and say, as most toddlers can, "chair"?

As computers have mastered rarefied domains once thought to be uniquely human, they simultaneously have failed to master the ground-floor basics of the human experience—spatial orientation, object recognition, natural language, adaptive goal-setting—and in so doing, have shown us how impressive, computationally and otherwise, such minute-to-minute fundamentals truly are.

We forget how impressive we are. Computers are reminding us.

One of my best friends was a barista in high school. Over the course of a day, she would make countless subtle adjustments to the espresso being made, to account for everything from the freshness of the beans to the temperature of the machine to the barometric pressure's effect on the steam volume, meanwhile manipulating the machine with an octopus's dexterity and bantering with all manner of customers on whatever topics came up. Then she went to college and landed her first "real" job: rigidly procedural data entry. She thought longingly back to her barista days—when her job actually made demands of her intelligence.

Perhaps the fetishization of analytical thinking, and the concomitant denigration of the creatural—that is, animal—and bodily aspects of life are two things we'd do well to leave behind. Perhaps at last, in the beginnings of an age of AI, we are starting to *center* ourselves again, after generations of living slightly to one side—the logical, left-hemisphere side. Add to this that humans' contempt for "soulless" animals, our unwillingness to think of ourselves as descended from our fellow "beasts," is now challenged on all fronts: growing secularism and empiricism, growing appreciation for the cognitive and behavioral abilities of organisms other than ourselves, and, not coincidentally, the entrance onto the scene of an entity with considerably less soul than we sense in a common chimpanzee or bonobo—in this way AI may even turn out to be a boon for animal rights.

Indeed, it's entirely possible that we've seen the high-water mark of our left-hemisphere bias. I think the return of a more balanced view of the brain and mind—and of human identity—is a good thing, one that brings with it a changing perspective on the sophistication of various tasks.

It's my belief that only experiencing and understanding *truly* disembodied cognition—only seeing the coldness and deadness and disconnectedness of something that really *does* deal in pure abstraction, divorced from sensory reality—can snap us out of it. Only this can bring us, quite literally, back to our senses.

In a 2006 article about the Turing Test, the Loebner Prize co-founder Robert Epstein writes, "One thing is certain: whereas the confederates in the competition will never get any smarter, the computers will." I agree with the latter, and couldn't disagree more strongly with the former.

When the world-champion chess player Garry Kasparov defeated Deep Blue, rather convincingly, in their first encounter in 1996, he and IBM readily agreed to return the next year for a rematch. When Deep Blue beat Kasparov (rather less convincingly) in '97, Kasparov proposed another rematch for '98, but IBM would have none of it. The company dismantled Deep Blue, which never played chess again.

The apparent implication is that—because technological evolution seems to occur so much faster than biological evolution (measured in years rather than millennia)—once the *Homo sapiens* species is overtaken, it won't be able to catch up. Simply put: the Turing Test, once passed, is passed forever. I don't buy it.

Rather, IBM's odd anxiousness to get out of Dodge after the '97 match suggests a kind of insecurity on its part that I think proves my point. The fact is, the human race got to where it is by being the most adaptive, flexible, innovative, and quick-learning species on the planet. We're not going to take defeat lying down.

No, I think that, while the first year that computers pass the Turing Test will certainly be a historic one, it will not mark the end of the story. Indeed, the *next* year's Turing Test will truly be the one to watch—the one where we humans, knocked to the canvas, must pull ourselves up; the one where we learn how to be better friends, artists, teachers, parents, lovers; the one where we come back. More human than ever.

This article available online at:

http://www.theatlantic.com/magazine/archive/2011/03/mind-vs-machine/308386/