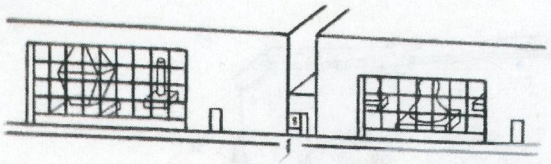
THE CRITICS



BOOKS

GET SMART

How will we know when machines are more intelligent than we are?

BY ADAM GOPNIK

When I was a small boy, there was a machine in the Franklin Institute, in Philadelphia, that played tic-tac-toe and never lost. No matter where you placed your X, it rebounded with the right O. It could always win or force a draw, even if you went first and took the center square. The machine looked smart to an eight-year-old, but my mother, a logician, linguist, and early Fortran speaker, explained to me on one of our frequent visits that smart was the *last* thing it was. It could do one thing—play one essentially dumb game—and it could do it only because it had been programmed to follow a mechanical network of on/off switches. It wasn't thinking; it was just tracking.

Now, if the machine could talk, she went on . . . well, then we'd be talking. She eventually became a researcher on an early, ambitious project in "machine translation," which aimed at creating a program for the Canadian government that would translate French into English and English into French, overnight, for Hansard, the parliamentary record. The dream was to write an algorithm, a sort of instant grammar, that the machine would learn and then use to translate sentences at warp speed. (Until then, a team of translators labored to do this, with a significant lag time.) Yet that simple-sounding task, done by bilingual Canadians each day, proved to be fiendishly, instructively, difficult to automate. The algorithm that connected a vocabulary term, seemingly transparent (*homme* = man) within a grammar, and seemingly easily mirrored (*l'homme est* = the man is; *l'homme était* = the man was), remained elusive. So many things that humans are naturally good at—guessing contexts, inferring rules, remembering oddities, resolving ambiguities, and, above all, grasping meaning from fragmentary "input"—defeated the programmers' efforts, leaving the machine tongue-tied.

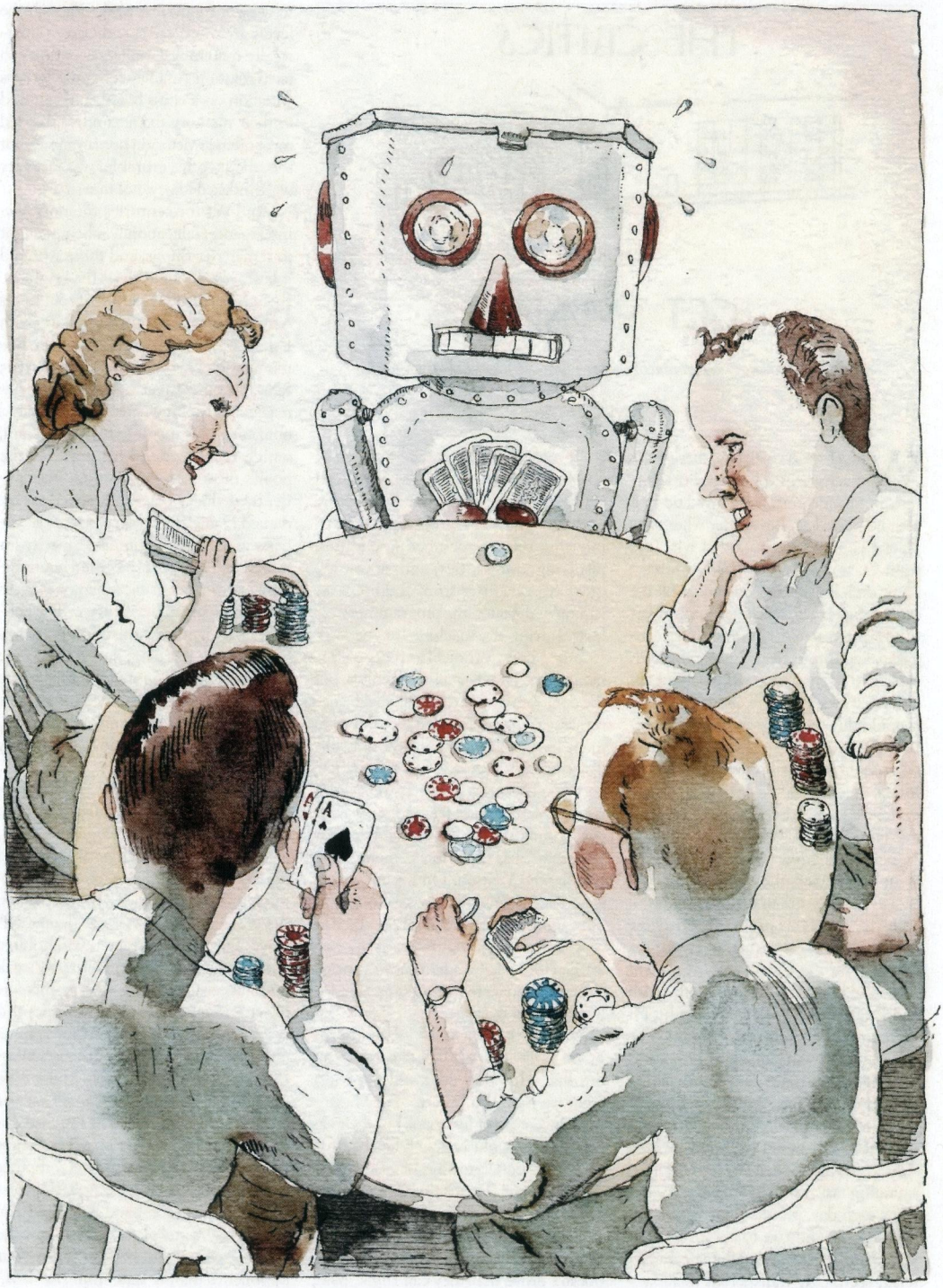
Since then, various frontiers of human smarts have fallen to computers, but the approach of true artificial intelligence remains as disputable and, according to many programming moms and dads, just as far off as it ever did. Tic-tac-toe fell, and then checkers, and chess went down when Deep Blue defeated Garry Kasparov, and now the TV trivia quiz show "Jeopardy!" has fallen to a computer system named Watson. On the surface, at least, what Raymond Kurzweil, in accents both ominous and worshipful, calls "the Singularity"—the "Matrix" moment when artificial intelligence becomes as strong as, if not stronger than, the human kind—gets closer all the time.

But are the new machines and programs really smarter? The skeptics point out that what they do is still not really what we mean by "smart." They have a huge inventory of instances, but their capacity for thought isn't so different from that of the tic-tac-toe machine in the science museum. They have big memories, and an amazing ability to race through them quickly to find the right thing to apply to the circumstance—but that doesn't prove that they can think, plan, strategize, surprise, or come up with a plan so crazy it might just work. Though the levels are more various, they're still basically matching familiar scenario A to fixed solution A. They recognize a chess situation on a chess board, and can find in their memory the next move that led most often to victory in human games, but this, the skeptics grumble, was just very well-indexed idiocy, not true smarts.

And yet for centuries memory *was* intelligence: education was being taught to remember things, and those who did it best were thought to be the brightest. By now, as Joshua Foer points out in his highly entertaining "Moonwalking with Einstein: The Art and Science of Remembering Everything" (Penguin Press; $26.95), our retreat from memory has reached the point where one scientist has outsourced his memory to a SenseCam, which hangs around his neck. In the book's most arresting section, Foer patiently deduces that a self-styled savant named Daniel Tammet, who dazzles experts with his computerlike capacity for total recall, is probably a gifted but traditional performer using mnemonic tricks older than the vaudeville stage. The ability to perform memorization feats was once in itself dazzling; now the feats are dazzling only if you pretend that you're not really doing them—memory amazes only if it's presented as a gift, not as a skill, since the skill is already, so to speak, fully mechanized.

In the same way, Watson, the "Jeopardy!" computer, as Stephen Baker explains, at intricate and eloquent length, in "Final Jeopardy: Man vs. Machine and the Quest to Know Everything" (Houghton Mifflin Harcourt; $24), was playing the game not like a normal "Jeopardy!" player but like a computer, with a huge database to draw on and a relatively weak internal censor against offering incredibly stupid responses when they popped up. The key to the treasure-house turns out to be the warehouse. Although Watson was vastly more suave and subtle than the old machines, it was still a server stuffed with answers and a retrieval device to find them. Organized dumbness is what beat human smarts. Computer programs are still revealingly hopeless at three- and four-hand poker, where, as they say, you "play the man, not the hand"—where you have to guess your opponent's mental state from his "tells," not his hole cards from your knowl-

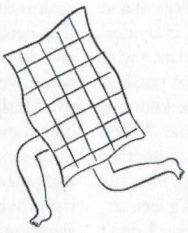
ABOVE: PHILIPPE WEISBECKER; OPPOSITE: BARRY BUTT



Computers are still revealingly hopeless at multi-hand poker—where you have to guess your opponent's mental state.

edge of the odds. Win at three-hand poker, with its three-dimensional mind-modelling . . . and then we'll talk.

Or, rather, really talk, and we won't have to worry about three-hand poker. If a program could consistently counterfeit human language in an ongoing exchange such that you couldn't tell that you were conversing with a computer, then, many theorists have argued, the threshold of intelligence would have been crossed, and there would be no need for more games to conquer. This is the famous "Turing test," named for Alan Turing, the tragic gay British genius who helped break the Enigma code, and who first proposed that the true test for artificial intelligence was to see if, cut off by a "black curtain," you would be able to tell a machine's answer from a human's in a natural-language conversation. Break that barrier, and the computer is as good as a guy or girl.



I had always thought of the Turing test as an abstract conceit, a philosopher's question, but it turns out to have produced actual tournaments—as though Zeno's paradox had produced real races run between tortoises and Greek warriors. The ins and outs of Turing tests and tournaments are the subject of the poet and computer aficionado Brian Christian's terrific "The Most Human Human" (Doubleday; $27.95), one of the rare successful literary offspring of "Gödel, Escher, Bach," where art and science meet an engaged mind and the friction produces real fire. Christian looks at a British Turing tournament, and its history, along with a parallel control tournament, to see if a human can seem human in a remote conversation: the "most human human" is challenged to shape his or her typing so that it unmistakably signals that it's a person talking.

Christian, in a book dense with ideas—ranging from an "anti-Lincoln-Douglas" debate format that every progressive school should take up to a discussion of why movie trailers are usually better than the movies they condense—points out that how competitors fare in the Turing tests is more about the style of the response than about the substance. Human intelligence expressed in sentences doesn't have only attributes and attainments; it has affect. Our stance, our emotional tone, is a surer sign that it's us back there than the ability to answer skill-testing questions. We interrupt, infer, guess, exclaim, ignore. And, at a deeper level, we express a "meta-attitude" about what we're saying and doing even as we say and do it. Human beings have the ability not only to win at "Jeopardy!" but to feel a little embarrassed about winning on "Jeopardy!" Nor is this affect merely "emotional," in the classic Kirk-to-Spock sense. ("Have you . . . no feelings! No ability. To go . . . beyond logic and know what's in . . . a human heart?") Empathy and sympathy, jokes and wordplay, are as necessary to intelligence as pure reason: the poker-playing program breaks down because it can't put itself in the mind of the guy across the table. Wit and puns aren't just décor in the mind; they're essential signs that the mind knows it's on, recognizes its own software, can spot the bugs in its own program.

Christian's central point is that the Turing bots that work best, whether produced by a computer or shaped by a mind, have to be, or fake being, dynamic. The best test of their humanness is not how smartly they offer answers but how quickly they interrupt, get distracted, compress information into slang codes, rely on "uh"s and "ah"s. Intelligence is an affect engaged in an *activity*. It flits between the empty spaces as much as it takes place in the exchanges. If a teen-age boy says to a teen-age girl, "I was, like, wondering, if, like, you'd like to, like, go to that, uh, thing at Jacob's?" and she says, "Uh, well . . ." it's bad news. But if she says, "Well, um . . ." it's promising, and if she says, "Yeah, like, funny, because, um . . ." it's the best news of all. Prefixes and tics and characteristic mannerisms are richly coded with information. The two best Presidential communicators of recent decades had distinctive vocal prefixes that did a lot of their talking for them: Ronald Reagan's "Well . . ." meant "Despite your attempt to antagonize me, I'm still going to appeal to plain old placid common sense," while Obama's "Look . . ." means "Forgive me if I sound impatient, but if you'll actually examine the facts in the case you'll see I'm right." One marker assures Capraesque cheer; the other, Spockian certitude. And it's hard to make either understandable to a machine.

Christian goes on to make the subtler, poetic point that human talk is not just an exchange of axioms, or even of emotionally coded abbreviations, but an activity played on an edge between the "lossiness" of compressed communication and the nimbleness with which we compress it—between our knowledge that in everything we say we have to leave out a lot of information for economy's sake and our ability to make that economy itself eloquent and informative. Kid-speak, again, is an ideal instance of compression in balance with concision. What sounds to the outsider limited and repetitive is to the knowing listener as nuanced as Henry James. When one eleven-year-old girl says to another eleven-year-old girl, "So then, like, the teacher got all, like, all of you, I guess, are, like, going to have to do a, like, I don't know, a makeup test. So! Like, *yeah*," she means: "The teacher, becoming heated"—that's why she "*got*, like," rather than "*said*, like"—"announced, in effect, that many of us (I suppose, at a first approximation, all) will, at some point in, as it were, the near future, have to take what actually amounts to, when all is said and done, a secondary makeup test. I have indignant feelings about this—as who among us would not?—but I recognize their essential futility." All of this is completely clear to the knowing listener, but it's been impossible, so far, to teach a machine to, you know, like, really, like, get it.

Yet human intelligence has another force, too: the sense of urgency that gives human smarts their drive. Perhaps our intelligence is not just ended by our mortality; to a great degree, it *is* our mortality. Imagine, for the moment, a set of networked, self-correcting computers, programmed to pursue a deliberately vague, long-term goal: say, "Make as many significant calculations as you can, and try to make more than any other computer in the lab," with "significant" left deliberately multivalent, open-ended. Then imagine that each of these computers has a stick of explosives by its C.P.U., with a slow-acting and temperamental seventy-year fuse, and that each knows it. And add that the corrosive acid that detonates the fuse also gradually slows down each computer's functioning—so that it's much better off, more likely to make more significant calculations, by interfacing with another computer before its own connections have worn down. The computers would, at every moment, have to make fiendishly difficult decisions about whether a particular calculation was worth investing in, given the larger, life-limited task of making truly significant calculations. They would have to calculate the benefits and losses, for instance, of intense exchange early on against the knowledge of their approaching annihilation, and against the demands of all the other tasks that needed doing. Some would retreat and do nothing but solitary calculations; some would network furiously; some would ask if it was worth trying to win TV quiz shows when the aim was to win the most-significant-calculations contest. The computers would make calculations of the right balance between time taken and significance achieved and distribute these within the network. Given the time pressures, the calculations would probably be short—say, ten or eleven lines—and the most insightful would probably be shared by all the other machines. (They might even be made more easily memorable by being set to rhythms and melodic patterns.) Some machines would doubtless start producing subprograms that meditated more abstractly on the difficulties of being a smart machine with an explosion waiting. ("At my back I always hear time's winged program drawing near"; "Gather your subroutines while you may.") Within a generation, irony, poetry, ambiguity, ecstasy would all be parts of the computers' output and affect. They would be smart and stupid as we are smart and stupid.

For the time being, the Singularity approaches, and we find ourselves at war not with an army of dying computers or super-intelligent robots who look like Daryl Hannah but with immobile, pathetic nerds who have somehow memorized the phone book and the encyclopedia. They have even made progress with my mom's problem. Stephen Baker explains that Google has approached the difficulty of machine translation not by coming up with a "Hansard algorithm" but by essentially entering *all of Hansard* into the program, creating such a huge corpus of French/English sentences that it can almost always find a near-match for any sentence you give it and tap-dance its way around the rest. It works, kind of.

But only kind of. A famous apologetic sentence from Albert Camus's Nobel Prize acceptance speech becomes, in Googleized English: "How can a man almost young, rich only in his doubts and still a work in progress, accustomed to live in solitude at work or pensions in friendship, should he not learned with a sort of panic which carried off at once, and only reduces to itself in the middle of a harsh light?" The most egregious error is typical of what still goes wrong: Camus uses the French word *retraites* in the less common sense of "retreats," "escapes"; the program has it stored in the more common French sense of "old-age pensions." A human speaker grasps instantly that "pensions" makes no sense in the context, but this isn't something the program is programmed to do.

Almost every sentence in Google Translate is tilted, weird, or just plain wrong—yet the general sense is tolerable, coming ever closer to the crossover moment when, by sheer statistical force, it does the job better than we can. Perhaps the real truth is this: the Singularity is not on its way—the Singularity happened long ago. We have been outsourcing our intelligence, and our humanity, to machines for centuries. They have long been faster, bigger, tougher, more deadly. Now they are much quicker at calculation and infinitely more adept at memory than we have ever been. And so now we decide that memory and calculation are not really part of mind. It's not just that we move the goalposts; we mock the machines' touchdowns as they spike the ball. We place the communicative element of language above the propositional and argumentative element, not because it matters more but because it's all that's left to us. The machines think, and we say that thinking is really the shrugs we make between statements; they talk, and we say that talk is really just the sounds we make between sentences. Doubtless, even as the bots strap us down to the pods and insert the tubes in our backs, we'll still be chuckling, condescendingly, "They look like they're thinking, sure, very impressive—but they don't have the affect, the style, you know, the vibe of real intelligence. It's just zombie stuff. Watch those buckles, cyber boy!" But by then we'll know, at least, the unique characteristic of human intelligence. What do we really mean by "smart"? The ability to continually diminish the area of what we mean by it. ♦