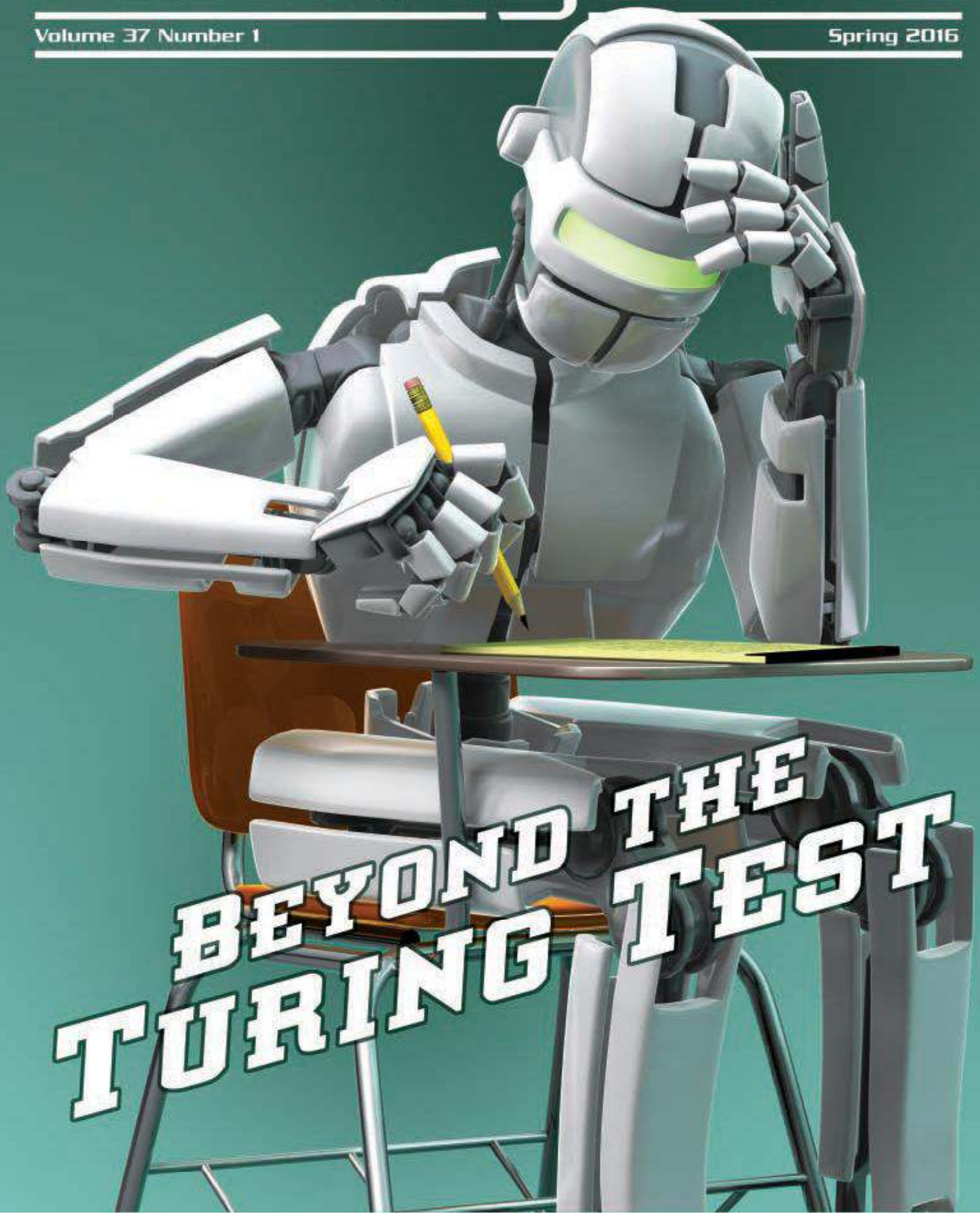
AI magazine
Volume 37 Number 1
Spring 2016



BEYOND THE TURING TEST

WWTS (What Would Turing Say?)

Douglas B. Lenat

■ *Turing's Imitation Game was a brilliant early proposed test of machine intelligence — one that is still compelling today, despite the fact that in the hindsight of all that we've learned in the intervening 65 years we can see the flaws in his original test. And our field needs a good "Is it AI yet?" test more than ever today, with so many of us spending our research time looking under the "shallow processing of big data" lamppost. If Turing were alive today, what sort of test might he propose?*

WTDS (What Turing Did/Didn't Say)

If you are reading these words, surely you are already familiar with the Imitation Game proposed by Alan Turing (1950). Or are you?

Turing was heavily influenced by the World War II "game" of allied and axis pilots and ground stations each trying to fool the enemy into thinking they were friendlies. So his imagined test for AI involved an interrogator being told that he or she was about to interview a man and woman over a teletype, both of whom would be pretending to be the woman; the task was to guess which one was lying. If a machine could fool interrogators as often as a typical man, then one would have to conclude that that machine, as programmed, was as intelligent as a person (well, as intelligent as *men*.)[1] As Judy Genova (1994) puts it, Turing's originally proposed game involves not a question of species, but one of gender.[2]

The current version, where the interrogator is told he or she needs to distinguish a person from a machine, is (1) much more difficult to get a program to pass, and (2) almost all the added difficulties are largely irrelevant to intelligence! And it's possible to muddy the waters even more by some programs appearing to do well at it due to various tricks, such as having the interviewee program claim to be a 13-year-old Ukrainian who doesn't speak English well (University of Reading 2014), and hence having all its wrong or bizarre responses excused due to cultural, age, or language issues.

Going into more detail here about why the current version of the Turing test is inadequate and distracting would be a

 ISSN 0738-4602

digression from my main point, so I've included that discussion as a sidebar to this article.

Here, let it suffice for me to point out that one improvement would be simply to go back to his originally proposed test, or some variant of it. I'm imagining here a game similar to the TV program *To Tell the Truth*. Panelists (the interrogators) are told that they are talking to three people who will all be claiming that some fact is true about them (for example, they treat sick whales; they ate their brother's bug collection; and others) and that two of the people are lying and one is telling the truth; their job is to ask questions to pick out the truth teller.

In my imagined game, the interrogator is told he or she will be interviewing three people online, all claiming *X*, and her or his task is to pick out the one truth teller. Then we measure whether our supposed AI fools the interrogator at least as often as the human "liars" are able to. Averaged over lots of interrogators, lots of claims, and lots of liars, this might be an improvement over today's Turing test.

Does that go far enough? It still smacks of a challenge one might craft for a magician. I can imagine programs doing well at that task through tricks, but then clearly (through subsequent failed attempts to apply them) revealing themselves not to be generally intelligent after all. So let's rethink the test from the top down.

WTMS (What Turing Might Say)

So what might Turing say today, if he were alive to propose a new test for machine intelligence? He was able to state the original test in one paragraph; he might first try to find an equally terse and compelling modern version.

Mathematics revolutionized physics in the late nineteenth and early twentieth centuries, and "softer" sciences like psychology and sociology and AI have been yearning not to be left behind. That type of physics envy has all too often led to premature formalization, holding back progress in AI at least as much as helping it. To quote economist Robert Heilbroner, "Mathematics has given economics rigor, but alas, also mortis."

I don't quite have enough presumption to claim that Turing would come up with the same test that I'm about to discuss, but I do believe that he'd recoil a bit at some of the tricks-based chatbots crafted in his name, and think twice before tossing off a new glib two-sentence-long test for AI.

My test, like his original Imitation Game, is one for recognizing AI when it's here. Instead of focusing on one computer program being examined for intelligence, what matters is that human beings synergizing with the AI exhibit what from our 2016 point of view would be superhuman intelligence.

The way to test for that, in turn, will be to look for the many and dramatic impacts that state of affairs would have on us, on our personal and professional lives, and on the way that various aspects of society and economy work. Some of the following are no doubt wrong, and will seem naïve and even humorous 65 years from now, but I'd be genuinely surprised[3] if *real AI* — from now on let's just call that RAI — didn't engender most of the following.

PDA

Almost everyone has a cradle-to-grave general personal assistant application that builds up an integrated model of the person's preferences, abilities, interests, modes of learning, idiosyncratic use of terms and expressions, experiences (to analogize to), goals, plans, beliefs. Siri and Cortana are indicators of how much demand there is for such PDAs. The real test for this having "arrived" will be not just its universal adoption but metalevel phenomena including legislation surrounding privacy and access by law enforcement; and the rise of standards and applications using those standards that broker communication between multiple individuals' PDAs; and marketing directed at the PDAs that will be making most of the mundane purchasing decisions in their ratava's (the inverse of "avatar") life.

Education

The popularity of massive open online courses (MOOCs) and the Khan Academy are early indicators of how much demand there is even for non-AI-based education courseware. When AI is here, we will see widespread individualized (using — and feeding back to — one's PDA) education to the point where in effect everyone is home schooled, "schools" continuing to exist in some form to meet the infrastructure, extracurricular, and social needs of the students. A return to what appears to be the monitorial system, where much of the student's time is spent emulating not so much a sponge (trying to absorb concepts and skills, as is true today) as emulating a teacher, a tutor, since — I think we've all experienced this — we often really understand something only after we've had to teach or explain it to someone else. In this case, the human (let's refer to her or him as the tutor) will be tutoring one or more tutees who will likely be AIs, not other human beings. Those "tutee" AIs will be constantly assessing the tutor and deciding what mistakes to make, what confusions to have, what apparent learning (and forgetting) to exhibit, based on what will best serve that tutor pedagogically, what will be motivated by situations in that person's real life (teaching you new things in situations where they would be useful and timely for you to know), based on the AI reasoning about what will be fun and entertaining to the person, and similar concerns that in effect blur the boundaries of what education is, compared with today.

Health Care

The previous two impacts ripple over to this — your PDA watching out for you and helping you become a more accurately and more fully informed consumer of health-care products and services, calling attention to things in ways and at times that will make a difference in your life. From the other direction, though, RAI will enable much more individualized diagnosis and treatment; for an early step along that line, see DARPA's Big Mechanism project, which has just begun, whose goal is to use AI to read and integrate large amounts of cancer research literature, which (coupled with patient-specific information) will enable plausible hypotheses to be formed about the pathways that your cancer is taking to grow and metastasize, and plausible treatments that might only be effective or even safe for you and a tiny sliver of other individuals. RAI (coupled with robotics only slightly more advanced than the current state of the art) will also revolutionize elderly care, given almost limitless patience, ability to recognize what their "patient"/companion is and isn't doing (for example, exercise-wise), and so on. This will later spread to nursing care for wider populations of patients. I fear that extending this all the way to child and infant care will be one of the last applications of AI in health care due to the public's and the media's intolerance of error in that activity.

Economy

This is currently based on atoms (goods), services involving atoms, and information treated as a commodity. The creation and curation of knowledge is, by contrast, done for free — given away in return for your exposure to online advertising and as a gateway to other products and services. I believe that RAI will change that, profoundly, and that people will not hesitate to be charged some tiny amount (a penny, let's say) for each useful alert, useful answer, useful suggestion. That in turn will fuel a knowledge economy in which contributors of knowledge are compensated in micropayment shares of that penny. Once this engine is jump-started, widespread vocation and avocation as knowledge contributors will become the norm. Some individuals will want and will receive the other sort of credit (citation credit) in addition or instead of monetary credit, possibly pseudonymously. Moreover, as we increase our trust in our PDA (above), it will be delegated increasing decision-making and spending authority; the old practice of items being sent to individuals "on approval" will return and human attention being paid to shopping may be relegated to hobby status, much as papermaking or home gardening today. Advertising will have to evolve or die, once consumers are better educated and increasingly the buying decisions are being made by their PDAs anyway. And ever-improving translation and (not using AI particularly) three-dimensional printing technologies will make the consumer's uncorrected physical location almost as unimportant as his or her uncorrected vision is today.

The flip side of the impact of AI on the economy is that a very small fraction of the population will be needed to grow the world's food and produce the world's goods, as robots reliably amplify the ability of a relatively few people to meet that worldwide demand. This will lead to something that many critics will no doubt label *universal socialism* in their then vastly greater free time.

Democracy and Government

RAI will probably have a dramatic effect in this area, pummeling the status quo of these institutions from multiple directions: for example, more effective education will result in a voting public better able to perform critical thinking and to detect and correct for attempts at manipulation and at revising history. Lawmakers and the public will be able to generate populations of plausible scenarios that enable them to better assess alternative proposed policies and courses of action. Fraud and malfeasance will become more and more difficult to carry out, with multiple independent AI watchdogs always awake and alert. Government functions currently drowning in red tape, due to attempts to be frugal through standardization, may be catalyzed or even automated by RAI, which can afford to — which will inevitably — know and treat everyone as an individual.

Our Personal Experience

By this I mean to include various sorts of phenomena that will go from unheard of to ubiquitous once RAI arrives. These include the following.

Weak Telepathy

You formulate an intent, and have barely started to make a gesture to act on it, when the AI understands what you have in mind and why, and completes that action (or a better one that accomplishes your actual goal) for you; think of an old married couple finishing each other's sentences, raised to the *n*th power. This isn't of course real telepathy — hence the word *weak* — but functionally is almost indistinguishable from it.

Weak Immortality

Your PDA's cradle-to-grave model of you is good enough that, even after your death, it can continue to interact with loved ones, friends, business associates, carry on conversations, carry our assigned tasks, and others; eventually this will be almost as though you never died (well, to everyone except you, of course, hence the word *weak*).

Weak "Cloning"

The quotation marks refer to the science-fiction type of duplication of you instantly as you are now, able to be in several places at once, attending to several things at once, with your one "real" biological consciousness and (through VR) awareness flitting to

The Current Turing Test Is Hard in Ways Both Unintended and Irrelevant

At AAAI 2006, I went through this at length (Lenat 2008), but the gist is that Turing's game had a human interrogator talking through a teletype with a man and a woman, both pretending that they were the woman. The experimenter measures what percentage of the time the average interrogator is wrong — identifies the wrong interviewee as being the woman. Turing's proposed test, then, was to see if a computer could be programmed to fool the interrogator (who was still told that they were talking to a human man and a human woman!) into guessing incorrectly about which interrogatee was the woman at least as often as men were able to fool the interrogator. One could argue then that such a computer, as programmed, was intelligent. Well, at least as intelligent the typical human male.[5]

Why is the revised gender-neutral version harder to pass and less reflective of human intelligence? If the interrogator is told that the task is to distinguish a computer from a person, then they can draw on his or her array of facts, experiences, visual and aural and olfactory and tactile capabilities, current events and history, expectations about how accurately and completely the average person remembers Shakespeare, and so on, to ask things they never would have asked under Turing's original test, when they thought they were trying to distinguish a human man from a human woman through a teletype.

Our vast storehouse of common sense also makes it more difficult to pass the "neutered" Turing test than the original version. Every time we see or hear a sentence with a pronoun, or an ambiguous word, we draw on that reservoir to decode what the author or speaker encoded into that shorthand. Most of the examples I've used in my talks and articles for the last 40 years (such as disambiguating the word pen in "the box is in the pen" versus "the pen is in the box") have been borrowed and reborrowed from Bar-Hillel, Chomsky, Schank, Winograd, Woods, and — surprisingly often and effectively — from Burns and Allen. Almost all of these disambiguatings are gender neutral — men perform them about as well as women perform them — hence they simply wouldn't come up or figure into the original Turing test, only the modern, neutered one.

The previous two paragraphs listed various ways in which the gender-neutral Turing test is made vastly more difficult because of human beings' gender-independent general knowledge and reasoning capabilities. The next few paragraphs list a few ways in which the gender-neutral Turing test is made more difficult because of gender-independent human foibles and limitations.

Human beings exhibit dozens of translogical behaviors: illogical but predictable wrong decisions that most people make, incorrect but predictable wrong answers to queries. Since they are so predictable, an interrogator in today's "neutered" Turing test could use these to separate human from nonhuman interrogatees, since that's what they are told their task is. As I said in 2008 (Lenat 2008): "Some of these are very obvious and heavy-handed, hence uninteresting, but still work a surprising fraction of the time — 'work' meaning, here, to enable the interrogator instantly to unmask many of the programs entered into a Turing test competition as programs and not human beings: slow and errorful typing; 7 +/– 2 short-term memory size; forgetting (for example, what day of the week was April 7, 1996? What day of the week was yesterday?); wrong answers to math problems (some wrong answers being more 'human' than others: 93 – 25 = 78 is more understandable than if the program pretends to get a wrong answer of 0 or –9998 for that subtraction problem. [Brown and van Lehn 1980]). … Asked to decide which is more likely, 'Fred S. just got lung cancer.' or 'Fred S. smokes and just got lung cancer,' most people say the latter. People worry more about dying in a hijacked flight than the drive to the airport. They see the 'face' on Mars. They hold onto a losing stock too long because of ego. If a choice is presented in terms of rewards, they opt for a different alternative than if it's presented in terms of risks. They are swayed by ads."

When faced with a difficult decision, human beings often select the alternative of inaction — if it is available to them — rather than action. One example of this is the startling statistic that in those European countries that ask driver's license applicants to "check this box to opt in" to organ donation, there is only a 15 percent enrollment, whereas in neighboring, culturally similar countries where the form says "check this box to opt out" there is an 85 percent organ donor enrollment. That is, 85 percent don't check the box no matter what it says! This isn't because this decision is beneath their notice, quite the contrary: they care very deeply about the issue, but they are ambivalent, and thus their reaction is to make the choice that doesn't require them to do anything, not even check a box on a piece of paper. Another, even more tragic, example of this "omission bias" (Ritov and Baron 1990) involves American parents' widespread reluctance to have their children vaccinated.

For more examples of these sorts of irrational yet predictable human behaviors, see, for example, Tversky and Kahneman (1983).

As an exercise, imagine that an extraterrestrial lands in Austin, Texas, and wants to find out how Microsoft Word works, the program I am currently running as I type these words. The alien carefully measures the cooling fan air outflow rate and temperature, and the disk-seeking sounds that my computer makes as I type these words, and then spends 65 years trying to mimic those air-heatings and clicking noises so precisely that no one can distinguish them from the sounds my Dell PC is making right now. Absurd! Pathetic! But isn't that in effect what the "neutered" Turing test proponents are requiring we do, requiring that our program do if it is to be adjudged to pass their test? Are we really so self-enthralled that we think it's wise to spend our precious collective AI research time getting programs to mimic the latency delays, error rates, limited short-term memory size, omission bias, and others, of human beings? Those aren't likely to be intimately tied up with intelligence, but rather just unfortunate artifacts of the platform on which human intelligence runs. They are about as relevant to intelligence as my Dell PC's cooling fan and disk noises are to understanding how Microsoft Word works.

whichever of your simulated selves needs you the most at that moment.

Arbitrarily Augmented Reality
This includes real-time correction for what is being said around and to you, so almost no one ever mishears or misunderstands any more. It includes superimposing useful details onto what you see, so you have the equivalent of X-ray and telescopic vision, and the sort of "important objects glow" effects seen in video games, paths of glowing particles to guide you, reformulation of objects you'd prefer to see differently (but with physical boundaries and edges preserved for safety).

Better-Than-Life Games and Entertainment
This is of course potentially dangerous and addictive, and — like many of the above predicted indicators — may herald very serious brand new problems, not just solutions to old ones.[4]

I'll close here, on that cautionary note. My purpose is not to provide answers, or even make predictions (though I seem to have done that), but rather to stimulate discussion about how we'll know when RAI has arrived: not through some Turing test Mark II but because the world will change almost overnight if or when superhuman aliens arrive — and *real AI* making its appearance is likely to be the one and only time that happens.

Acknowledgments

The following individuals provided comments and suggestions that have helped make this article more accurate and on-point, but they should not be held accountable for any remaining inaccuracies, omissions, commissions, or inflammations: Paul Cohen, Ed Feigenbaum, Elaine Kant, Elaine Rich, and Mary Shepherd.

Notes

1. Creepily, many people today in effect play this game online every day: men trying to "crash" women-only chats and forums, pedophiles pretending to be 10 year olds, MMO players lying about their gender or age, and others.

2. There remains some ambiguity (given his dialogue examples) about what Turing was proposing. But there is no ambiguity in the fact that the gender-neutral version is how the world came to recall what Turing wrote, by the time of the 1956 Dartmouth AI Summer Project, and ever since.

3. Alan Kay says that the best way to predict the future is to invent it. In that sense, these "predictions" could be recast as challenge problems for AI, a point of view consonant with Feigenbaum (2003) and Cohen (2006).

4. For example, while most of us will use AI to help us see multiple sides of an issue, to see reality more accurately and completely, AI could also be used for the opposite purpose, to filter out parts of the world that disagree with how we want to believe it to be.

5. He then gives some dialogue examples that make his intent somewhat ambiguous, but after that he returns to his main point about the computer pretending to be a man; and then discusses various possible objections to a computer ever being considered intelligent.

References

Brown, J. S., and VanLehn, K. 1980. Repair Theory: A Generative Theory of Bugs in Procedural Skills. *Cognitive Science* 4(4): 379–426.

Cohen, P. 2006. If Not Turing's Test, Then What? *AI Magazine* 26(4): 61–67.

Feigenbaum, E. A. 2003. Some Challenges and Grand Challenges for Computational Intelligence. *Journal of the Association for Computing Machinery* 50(1): 32–40.

Genova, J. 1994. Turing's Sexual Guessing Game. *Journal of Social Epistemology* 8(4): 313–326.

Lenat, D. B. 2008. The Voice of the Turtle: Whatever Happened to AI? *AI Magazine* 29(2): 11–22.

Ritov, I., and Baron, J. 1990. Reluctance to Vaccinate: Omission Bias and Ambiguity. *Journal of Behavioral Decision Making* 3(4): 263–277.

Turing, A. M. 1950. Computing Machinery and Intelligence. *Mind* 59(236): 433–460.

Tversky, A., and Kahneman, D. 1983. Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. *Psychological Review* 90(4): 293–315.

University of Reading. 2014. Turing Test Success Marks Milestone in Computing History. Press Release, June 8, 2014. Communications Office, University of Reading, Reading, UK (www.reading.ac.uk/news-and-events/releases/PR583836.aspx).

Doug Lenat, a prolific author and pioneer in artificial intelligence, focuses on applying large amounts of structured knowledge to information management tasks. As the head of Cycorp, Lenat leads groundbreaking research in software technologies, including the formalization of common sense, the semantic integration of — and efficient inference over — massive information sources, the use of explicit contexts to represent and reason with inconsistent knowledge, and the use of existing structured knowledge to guide and strengthen the results of automated information extraction from unstructured sources. He has worked in diverse parts of AI — natural language understanding and generation, automatic program synthesis, expert systems, machine learning, and so on — for more than 40 years now. His 1976 Stanford Ph.D. dissertation, AM, demonstrated that creative discoveries in mathematics could be produced by a computer program (a theorem proposer, rather than a theorem prover) guided by a corpus of hundreds of heuristic rules for deciding which experiments to perform and judging "interestingness" of their outcomes. That work earned him the IJCAI Computers and Thought Award and sparked a renaissance in machine-learning research. Lenat was on the computer science faculties at Carnegie Mellon University and Stanford, was one of the founders of Teknowledge, and was in the first batch of AAAI Fellows. He worked with Bill Gates and Nathan Myhrvold to launch Microsoft Research Labs, and to this day he remains the only person to have served on the technical advisory boards of both Apple and Microsoft. He is on the technical advisory board of TTI Vanguard, and his interest and experience in national security has led him to regularly consult for several U.S. agencies and the White House.