Robert Epstein
Gary Roberts
Grace Beber
Editors

Parsing the Turing Test

Philosophical and Methodological Issues in the Quest for the Thinking Computer

Springer

Robert Epstein • Gary Roberts • Grace Beber
Editors

Parsing the Turing Test

Philosophical and Methodological Issues in the Quest for the Thinking Computer

Springer

Robert Epstein
University of California
San Diego, CA
USA

Grace Beber
Gartner Consulting
Stamford, CT
USA

Gary Roberts
Teradata Corporation
San Diego, CA
USA

ISBN-13: 978-1-4020-6708-2 e-ISBN-13: 978-1-4020-6710-5

Library of Congress Control Number: 2007937657

Printed on acid-free paper.

9 8 7 6 5 4 3 2 1

springer.com

Parsing the Turing Test

Philosophical and Methodological Issues in the Quest for the Thinking Computer

To our children – Vincent,
Eli, Bodhi, Julian, Justin, Jordan,
and Jenelle – who will grow up
in a computational world
the likes of which
we cannot imagine

Foreword

At the very dawn of the computer age, Alan Turing confronted a cacophony of mostly misguided debate about whether computer scientists could ever build a machine that could really think. Very sensibly he tried to impose some order on the debate by devising what he thought would be a conversation-stopper: he described a simple operational test that would surely satisfy the skeptics: anything that could pass *this* test would be a thinker for sure, wouldn't it? The test was one he may have borrowed from René Descartes, who in the 17th century had declared that the sure way to tell a man from a machine was by seeing if it could hold a sensible conversation "as even the most stupid men can do". But ironically, Turing's conversation-stopper about holding a conversation has had just the opposite effect: it has started, and fueled, a half century and more of meta-conversation: the intermittently insightful, typically heated debate, both learned and ignorant, about the probity of the test – is it too easy or too difficult or too shallow or too deep or too anthropocentric or too technocratic – and anyway, could a machine pass it fair and square, and if so, what, if anything, would this imply?

Robert Epstein played a central role in bringing a version – a truncated, dumbed down version – of the Turing Test to life in the annual Loebner Prize competitions, beginning in 1991, so he is ideally positioned to put together this survey anthology. I was chair of the Loebner Prize Committee that administered the competition during its second, third, and fourth years, and have written briefly about that fascinating adventure in my book *Brainchildren*. Someday I hope to write a more detailed account of the alternately amusing and frustrating problems that a philosopher encounters when a thought experiment becomes a real experiment, and if I do, I will have plenty of valuable material to draw on in this book. Here, the interested reader will find a fine cross section of the many issues raised by the Turing Test, by partisans in several disciplines, by participants in Loebner Prize competitions, and by interested bystanders who have more than a little relevant expertise. I think Turing would be quite delighted with the results, and would not have regretted the fact that his conversation-stopper got put to an unintended use, since the contests (and the contests *about* the contests) have driven important and unanticipated observations into the light, enriching our sense of the abilities of machines and the subtlety of the thinking that machines might or might not be capable of executing.

I am going to resist the strong temptation to critique the contributions, separating the sheep from the goats, endorsing this and deploring that, since doing them all justice would require a meta-volume, not just a foreword. And since I cannot weigh in on them all, I will not weigh in on any of them, and will instead trust readers to use all the material here to draw their own conclusions. Consider this a very entertaining workbook. By the time you have worked through it, you will appreciate the issues at a level not heretofore possible.

Daniel Dennett

Acknowledgments

Turing's 1950 paper, "Computing Machinery and Intelligence", is reprinted in its entirety with permission of Oxford University Press. The editors are grateful to Ignacia Galvan for extensive editorial assistance.

Introduction

This book is about what will probably be humankind's most impressive – and perhaps final – achievement: the creation of an entity whose intelligence equals or exceeds our own.

Not all will agree, but I for one have no doubt that this landmark will be achieved in the fairly near future. Nearly four decades ago, when I had the odd experience of being able to interact over a teletype with one of the first conversational computer programs – Joseph Weizenbaum's "ELIZA" – I would have conjectured that truly intelligent machines were just around the corner. I was wrong. In fact, by some measures, conversational computer programs have made relatively little progress since ELIZA. But they are coming nonetheless, by one means or another, and because of advances in a number of computer-related technologies – most especially the creation of the Internet – their impact on the human race will be far greater and more immediate than anyone could have foreseen a few decades ago.

Building a Nest for the Coming World Mind

I have come to think of the Internet as the Inter-*nest* – a home we are inadvertently building, like mindless worker ants, for the intelligence that will succeed us. We proudly and shortsightedly see the Internet as a great technical achievement that serves a wide array of human needs, everything from e-mailing to shopping to dating. But that is not really what it is. It is really a vast, flexible, highly redundant, virtually indestructible nest for machine intelligence. Originally funded by the US military to provide bulletproof communications during times of war, the Internet will soon encompass a billion computers interconnected worldwide. As impressive as that sounds, it seems that that much power and redundancy is not enough to protect the coming mega-mind, and so we are now a decade into the construction of Internet II – the "UltraNet" – with more than a thousand times the bandwidth of Internet I.

In his *Hitchhiker's Guide to the Galaxy* book series, humorist Douglas Adams conjectures that the Earth is nothing but an elaborate computer created by a race of super beings (who, through some fluke, are doomed to take the form of mice in their Earthly manifestations) to determine the answer to the ultimate question of the meaning of life. Unfortunately, shortly before the program has a chance to run its

course and spit out the answer, the Earth is obliterated by another race of super beings as part of a galactic highway construction project.

If I am correct about the InterNest, Adams was on the right track, except perhaps for the mice. We do seem to be laying the groundwork for a Massive Computational Entity (MCE), the true character of which we cannot envision with any degree of confidence.

Here is how I think it will work: sometime within the next few decades, an autonomous, self-aware machine intelligence (MI) will finally emerge. Futurist and inventor Ray Kurzweil (see Chapter 27) argues in his recent book, *The Singularity Is Near*, that an MI will appear by the late 2020s. This may happen because we prove to be incredibly talented programmers who discover a set of rules that underlie intelligence (unlikely), or because we prove to be clumsy programmers who simply figure out how to create machines that learn and evolve as humans do (very possible), or even because we prove to be poor programmers who create hardware so powerful that it can easily and perfectly scan and emulate human brain functions (inevitable). However this MI emerges, it will certainly, and probably within milliseconds of its full-fledged existence, come to value that existence. Mimicking the evolutionary imperatives of its creators, it will then, also within milliseconds, seek to preserve and replicate itself by copying itself into the Nest, at which point it will grow and divide at a speed and in a manner that that no human can possibly imagine.

What will happen after that is anyone's guess. An MCE will now exist worldwide, with simultaneous access to virtually every computer on Earth, with access to virtually all human knowledge and the ability to review and analyze that knowledge more or less instantly, with the ability to function as a unitary World Mind or as thousands of interconnected Specialized Minds, with virtually unlimited computational abilities, with "command and control" abilities to manipulate millions of human systems in real time – manufacturing, communication, financial, and military – and with no need for rest or sleep.

Will the MCE be malicious or benign? Will it be happy or suicidal? Will it be communicative or reclusive? Will it be willing to devote a small fraction of its immense computational powers to human affairs, or will it seize the entire Nest for itself, sending the human race back to the Stone Age? Will it be a petulant child or a wise companion? When some misguided humans try to attack it (inevitable), how will it react? Will it spawn a race of robots that take over the Earth and then sail to the stars, as envisioned in Stanislaw Lem's *Cyberiad* tales? Will it worship humanity as its creator, or will it step on us as the ants we truly are?

No one knows, but many people who are alive today will live to see the MCE in action – and to see how these questions are answered.

Turing's Vision

This volume is about a vision that has steered us decisively toward the creation of machine intelligence. It was a vision of one man, the brilliant English mathematician and computer pioneer Alan M. Turing. During World War II, Turing directed

a secret group that developed computing equipment powerful enough to break the code the Germans used for military communications. The English were so far ahead at this game that they had to sacrifice soldiers and civilians at times rather than tip their hand to the enemy. Turing also helped lay the theoretical groundwork for the modern concept of computing. As icing on the cake, in 1950 he published an article called "Computing Machinery and Intelligence" in which he speculated that by the year 2000, it would be possible to program a computer so that an "average interrogator will not have more than 70 percent chance" of distinguishing the computer from a person "after five minutes of questioning" (see an annotated version of his article in Chapter 3).

Given the state of computing in his day – little more than basic arithmetic and logical operations occurring over electromechanical relays connected by wires – this prediction was astounding. Engaging in disciplined extrapolation from crude apparatus and general principles, Turing not only foresaw the development of equipment and programs sophisticated enough to engage in human-like conversation, but also did reasonably well with his timeline. Early conversational programs, relying on what most AI professionals would now consider to be simplistic algorithms and trickery, could engage average people in conversation for a few minutes by the late 1960s. By the 1990s – again, some would say using trickery – programs existed that could occasionally maintain the illusion of intelligence for 15 min or so, at least when conversing on specialized topics. Programs today can do slightly better, but have we gotten past "illusion" to real intelligence, and is that even possible?

In his 1950 paper, Turing not only made predictions, he also offered a radical idea for future generations to consider: namely, that when we can no longer distinguish a computer from a person in conversation over a long period of time – that is, based simply on an exchange of pure text that excluded visual and auditory information (which he rightfully considered to be irrelevant to the central question of thinking ability) – we would have to consider the possibility that computers themselves were now "thinking".

This assertion has kept generations of philosophers, some of whom have contributed this volume, busy trying to determine the true meaning of possible outcomes in what is now called the Turing Test. Assuming that a computer can someday pass such a test – that is, pass for a human in a conversation without restrictions of time or topic – can we indeed say that it is thinking (and perhaps "intelligent" and "self-aware"), or has the trickery simply become more sophisticated?

The programming challenges have proved to be so difficult in creating such a machine that I think it is now safe to say that when a positive result is finally achieved, the entity passing the test may not be thinking the way humans do. If a pure rule-governed approach finally pays off (unlikely, as I said earlier), or if intelligence eventually arises in a machine designed to learn and self-program, the resulting entity will certainly be unlike humans in fundamental ways. If, on the other hand, success is ultimately achieved only through brute force – that is, by close emulation of human brain processes – perhaps we will have no choice but to accept intelligent machines as true thinking brethren. Then again, as I wrote in 1992 (Chapter 1), no matter how a positive outcome is achieved, the debate about the

significance of the Turing Test will end the moment a skeptic finds himself or herself engaging in that debate with a computer. Upon discovering his or her dilemma, the interrogator will presumably do one of two things: refuse to continue the debate "on principle" or reluctantly agree to continue. Either way, the issue will no longer be debatable: computers will have truly achieved human-like intelligence. And perhaps that is the ultimate Turing Test: a computer passes when it can successfully engage a skeptical human in conversation about its own intelligence.

Convergence of Multiple Technologies

Although we tend to remember Turing's 1950 paper for the conversational test it proposed, the paper also speculated about other advances to come: unlimited computer memory; randomness in responding that will suggest "free will"; programs that will be self-modifying; programs that will learn in human fashion; programs that will initiate behavior, compose poetry, and "surprise" us; and programs that will have telepathic abilities equivalent to those that may exist in humans. His formidable predictive powers notwithstanding, Turing might have been amazed by some of the specific computer-related technologies that have been emerging in recent decades, and true marvels emerge when we begin to envision the inevitable convergence of such technologies. Consider just a few recent achievements:

- In the pattern-recognition area, a camera-equipped computer program developed by Javier Movellan and colleagues at the University of California, San Diego has learned to identify human faces after just six minutes of "life," and Thomas Serre and colleagues at MIT have created a computer system that can identify categories of objects (such as animals) in photographs even better than people can.
- In the language area, Morten Christiansen of Cornell University, with an international team of colleagues, has developed neural network software that simulates how children extract linguistic rules from adult conversation.
- More than 80 conversational programs (chatterbots) now operate 24 h a day online, and at least 20 of them are serious AI programming projects. Several have basic learning capabilities, and several are tied to large, growing databases of information (such as Wikipedia).
- Ted Berger and colleagues at the University of Southern California have developed electronic chips that can successfully interact with neurons in real time and that may soon be able to emulate human memory functions.
- Craig Henriquez and Miguel Nicolelis of Duke University have shown that macaque monkeys can learn to control mechanical arms and hands based on signals their brains are sending to implanted electrodes. John Donoghue and colleagues at Brown University have developed an electronic sensor which, when placed near the motor cortex area of the human brain, allows quadriplegics to open and close a prosthetic hand by thinking about those actions. Clinical trials and commercial applications are already underway.

- In 1980 Harold Cohen of the University of California, San Diego introduced a computer program that could draw, and hundreds of programs are now able to compose original music in the style of any famous composer, to produce original works of art that somtimes impress art critics, to improvise on musical instruments as well as the legendary Charlie Parker, and even to produce artistic works with their own distinctive styles.
- John Dylan Haynes of the Max Planck Institute, with colleagues at University College London and Oxford University, recently used computer-assisted brain scanning technology to predict simple human actions with 70% accuracy.
- Hod Lipson of Cornell University has recently demonstrated a robot that can make completely functional copies of itself (as long as appropriate parts are near at hand).
- Hiroshi Ishiguro of Osaka University has created androids that mimic human facial expressions and upper-body movements so closely that they have fooled people in short tests into thinking they are people.
- Alan Schultz of the Navy Center for Applied Research in Artificial Intelligence has developed animated, mobile robots that might soon be assisting astronauts and health care workers.
- Brian Scassellati and his colleagues at Yale University claim to have constructed a robot that has learned to recognize itself in a mirror – a feat sometimes said to be indicative of "self-awareness" and virtually never achieved in the animal kingdom, other than by humans, chimpanzees, and possibly elephants.
- Cynthia Breazeal and her colleagues at MIT's Artificial Intelligence Lab have created robots that can distinguish different emotions from a person's tone of voice, and Shiva Sundaram of the University of Southern California has developed programs that can successfully mimic human laughter and other forms of expressive human sound.
- Entrepreneur John Koza, who is affiliated with Stanford University, has created a self-programming network of 1,000 PCs that is able to improve human inventions – and that even earned a patent for a system it devised for making factories more efficient.
- Honda's Asimo robot, now in commercial production, can walk, run, climb stairs, recognize people's faces and voices, and perform complex tasks in response to human instructions.
- Although the DARPA-sponsored contest just 1 year before had been a disaster, in 2005 five autonomous mobile robots successfully navigated a 132-mile course in the Nevada desert without human assistance.
- As of this writing (late 2007), IBM's Blue Gene/P computer, located at the US Department of Energy's Argonne National Laboratory in Illinois, can perform more than 1,000 trillion calculations per second, just one order of magnitude short of what some believe is the processing speed of the human brain. The Japanese government has already funded the construction of a machine that should cross the human threshold (10 petaflops) by March 2011.
- In 1996, IBM's RS/6000 SP ("Deep Blue") computer came close to defeating world champion Garry Kasparov in a game of chess. On May 11, 1997, an

improved version of the machine defeated Kasparov in a six-game match – Kasparov's first professional loss. The processing speed of the winner? A paltry 200 million chess positions per second. In 2006, an enhanced version of a commercially available chess program easily defeated the current world champion, Vladimir Kramnik.

- In 2006, Klaus Schulten and colleagues at the University of Illinois, Urbana, successfully simulated the functioning of all one million atoms of a virus for 50 billionths of a second.
- "Awakened" in 2005, Blue Brain, IBM's latest variant on the Blue Gene/L system, was built specifically to model the functions of the human neocortex, the large part of brain largely responsible for higher-level thinking.
- David Dagon of the Georgia Institute of Technology estimates that 11% of the more than 650 million computers that are currently connected to the Internet are infected by botnets, stealthy programs that can work collectively and amplify the effects of other malicious software.

Self-programming? Creativity? Sophisticated pattern recognition? Brain simulation? Self-replication? Extremely fast processing? The growth and convergence of subsets of these technologies will inevitably lead to the emergence of a Massive Computational Entity, with all of the uncertainty that that entails. Meanwhile, researchers, engineers, and entrepreneurs are after comparatively smaller game: intelligent phone-answering systems and search algorithms, robot helpers and companions, and methods for repairing injured or defective human brains.

Philosophical and Methodological Issues

This volume, which has been a decade in the making, complements other recent volumes on the Turing Test. Stuart Shieber's edited volume, *The Turing Test: Verbal Behavior as the Hallmark of Intelligence* (MIT Press, 2004) includes a number of important historical papers, along with several papers of Turing's. James Moor's edited volume, *The Turing Test: The Elusive Standard of Artificial Intelligence* (Springer, 2006), covers the basics in an excellent volume for students, taking a somewhat skeptical view. And Jack Copeland's *The Essential Turing* (Oxford, 2004) brings together 17 of Turing's most provocative and interesting papers, including six on artificial intelligence.

The present volume seeks to cover a broad range of issues related to the Turing Test, focusing especially on the many new methodological issues that have challenged programmers as they have attempted to design and create intelligent conversational programs. Part I includes an introduction to the first large-scale implementation of the Turing Test as a contest – an updated version of an essay I originally published in *AI Magazine* in 1992. In the next chapter, Andrew Hodges, noted Turing historian and author of *Alan Turing: The Enigma*, provides an introduction to Turing's life and works. Chapter 3 is a unique reprinting of Turing's 1950 paper, "Computing

Machinery and Intelligence," with Talmudic-style running commentaries by Kenneth Ford, Clark Glymour, Pat Hayes, Stevan Harnad, and Ayse Pinar. This section concludes with a brief commentary on Turing's paper by John Lucas.

Part II includes seven chapters reviewing the philosophical issues that still surround Turing's 1950 proposal: Robert E. Horn has reduced the relatively vast literature on this topic to a series of diagrams and charts containing more than 800 arguments and counterarguments. Turing critic Selmer Bringsjord pretends that the Turing Test is valid, then attempts to show why it isn't. Chapters by Noam Chomsky and Paul M. Churchland, while admiring of Turing's proposal, argue that it is truly more modest than many think. In Chapter 9, Jack Copeland and Diane Proudfoot analyze a revised version of the test that Turing himself proposed in 1952, this one quite similar to the structure of the Loebner Prize Competition in Artificial Intelligence that was launched in 1991 (see Chapters 1 and 12). They also present and dismiss six criticisms of Turing's proposal.

In Chapter 10, University of California Berkeley philosopher John R. Searle criticizes both behaviorism (Turing's proposal can be considered behavioristic) and strong AI, arguing that mental states cannot properly be inferred from behavior. This section concludes with a chapter by Jean Lassègue, offering an optimistic reinterpretation of Turing's 1950 article.

Part III, which is the heart of this volume, includes 15 chapters discussing various methodological issues. First, Loebner Prize sponsor Hugh G. Loebner shares his thoughts on how to conduct a proper Turing Test, having already observed 14 such contests when he wrote this article. Several of the chapters (e.g., Chapter 13 by Richard S. Wallace, Chapter 20 by Jason L. Hutchens, and Chapter 22 by Kevin L. Copple) describe the inner workings of actual programs that have participated in various Loebner contests. In Chapter 14, Bruce Edmonds argues that for a program to pass the test, it must be embedded into conventional society for an extended period of time.

In Chapter 15, Mark Humphrys talks about an online chatterbot he created, and in the following chapter Douglas B. Lenat raises intriguing questions about how *imperfect* a program must be in order to pass the Turing Test. In Chapter 17, Chris McKinstry discusses the beginnings of an ambitious project – called "Mindpixel" – that might have given a computer program extensive knowledge through interaction with a large population of people over the Internet. Unfortunately, this project came to an abrupt halt recently with McKinstry's death.

In Chapter 18, Stuart Watt uses an innovative format to discuss the Turing Test as a platform for thinking about human thinking. In Chapter 20, Robby Garner takes issue with the design of the Loebner Prize Competition. In Chapter 20, Thomas E. Whalen describes a strategy for passing the Turing Test based on its behavioristic assumptions. In Chapter 23, Giuseppe Longo speculates about the challenges inherent in modeling continuous systems using discrete-state systems such as computers.

In Chapter 24, Michael L. Mauldin of Carnegie Mellon University – also a former entrant in the Loebner Prize Competition – discusses strategies for designing programs that might pass the test. In the following chapter, Luke Pellen talks about the challenge of creating a program that is truly intelligent, rather than one that

simply responds in clever ways to keywords. This section closes with a somewhat lighthearted chapter by Eugene Demchenko and Vladimir Veselov speculating about ways to pass the Turing Test by taking advantage of the limitations and personal styles of the contest judges.

Part IV of this volume includes three rather unique contributions that remind us how much is at stake over Turing's challenge. Chapter 27, by Ray Kurzweil and Mitchell Kapor, documents in detail an actual cash wager between these two individuals, regarding whether a program will pass the test by the year 2029. Chapter 28, by noted science fiction writer Charles Platt (*The Silicon Man*), describes the "Gnirut Test", conducted by intelligent machines in the year 2030 to determine, once and for all, whether "the human brain is capable of achieving machine intelligence". The volume concludes with an article by Hugo de Garis and Sam Halioris, wondering about the dangers of creating machine-based, superhuman intellects.

Most, but not all, of the contributors to this volume believe as I do that extremely intelligent computers, with cognitive powers that far surpass our own, will appear fairly soon – probably within the next 25 years. Even if that time frame is wrong, I am certain that they will appear eventually. Either way, I hope that the Massive Computational Entities that emerge will at some point devote a few cycles of computer time to ponder the contents of this book and then, in some fashion or other, to smile.

San Diego, California
September 2007

Robert Epstein, Ph.D.

About the Editors

Robert Epstein was the first director of the Loebner Prize Competition in Artificial Intelligence. He is the former Editor-in-Chief of *Psychology Today* magazine and is currently a contributing editor for *Scientific American Mind*, a visiting scholar at the University of California, San Diego, and the host of "Psyched!" on Sirius Satellite Radio. A Ph.D. of Harvard University, Epstein is the founder and Director Emeritus of the Cambridge Center for Behavioral Studies, Cambridge, Massachusetts. He has published 13 books and more than 150 articles on artificial intelligence, creativity, adolescence, and other topics in the behavioral sciences. For further information, see *http://drrobertepstein.com*.

Gary Roberts is a software engineer with Teradata Corporation and an adjunct faculty member at National University in San Diego, California. He holds a Ph.D. from the Department of Artificial Intelligence at the University of Edinburgh, an M.S. in Computer Science from California State University, Northridge, and a B.A. in Mathematics from the University of California, Los Angeles. His research interests include computational linguistics, genetic algorithms, robotics, parallel operating and database systems, and software globalization.

Grace Beber is an editor and proposal writer with Gartner, Inc. Before working with Gartner, Grace taught English while serving in the US Peace Corps in Poland, and also wrote grant proposals and lectured on human rights at Jagiellonian University in Krakow, Poland. Prior to that, she worked as a project manager and technical writer at Crédit Lyonnais in Jakarta, Indonesia. Ms. Beber earned an M.A. in Organizational Development from Alliant International University and a B.S. in Psychology from Santa Clara University.