Can a Computer Think?

Jeffrey Shallit
Department of Computer Science
University of Waterloo
Waterloo, Ontario N2L 3G1
Canada
shallit@graceland.uwaterloo.ca
http://www.math.uwaterloo.ca/~shallit

> "... a common form of scientific aberration, namely, the tendency of computer experts to be pontifical about subjects in which they have no competence."
>
> – Mortimer Taube, *Computers and Common Sense: The Myth of Thinking Machines*

Can a Computer Think?

What is meant by

- thought
  - *Can a computer think?*
- intelligence
  - *Can computers be intelligent?*
- consciousness
  - *Can computers be conscious?*
- free will
  - *Can computers possess free will?*
- belief
  - *Can machines have beliefs?*
- understanding
  - *Does a chess-playing program understand chess?*

Can a Computer Think?

"The ... question, 'Can machines think?' I believe to be too meaningless to deserve discussion. Nevertheless I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted."

– Alan Turing, 1950

The Turing Test

In 1950, computer scientist Alan Turing published a very influential paper in the journal *Mind* entitled, Computing Machinery and Intelligence.

VOL. LIX. No. 236.] [October, 1950

MIND

A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

BY A. M. TURING

In it, he proposed what is now known as the Turing test.

The Turing Test: Modern Version

- An interrogator communicates via a computer terminal with either a computer or person in the next room.
- Via a series of questions, the interrogator attempts to determine the identity of the correspondent in the next room: computer or person.
- A computer is said to pass the test if the interrogator cannot distinguish between computer and human with better than 50% probability.

Turing's Sample Dialogue

Q. Please write me a sonnet on the subject of the Forth Bridge.

A. Count me out on this one. I could never write poetry.

Q. Add 34957 to 70764.

A. (Pause about 30 seconds then give as an answer) 105621.

Q. Do you play chess?

A. Yes.

Q. I have my K at K1, and no other pieces. You have only K at K6 and R at R1. It is your move. What do you play?

A. (After a pause of 15 secs.) R-R8 mate.

The Turing Test

"On the other hand, if there were machines which bore a resemblance to our body and imitated our actions as far as it was morally possible to do so, we should always have two very certain tests by which to recognize that, for all that, they were not real men. The first is, that they could never use speech or other signs as we do when placing our thoughts on record for the benefit of others... And the second difference is, that although machines can perform certain things as well as or perhaps better than any of us can do, they infallibly fall short in others, by the which means we may discover that they did not act from knowledge, but only from the disposition of their organs. For while reason is a universal instrument which can serve for all contingencies, these organs have need of some special adaptation for every particular action. From this it follows that it is morally impossible that there should be sufficient diversity in any machine to allow it to act in all the events of life in the same way as our reason causes us to act."

– Descartes, Discourse on Method

The Turing Test

> "I believe that in about fifty years' time it will be possible to programme computers, with a storage capacity of about 10^9, to make them play the imitation game so well that an average interrogator will not have more than 70 per cent chance of making the right identification after five minutes of questioning."
>
> – Alan Turing, 1950

Hugh Loebner funded a US $100,000 prize for the first computer that can pass the test, and annual contests with smaller prizes.

The winner in 2000 was Alice by Richard Wallace.

Sample Dialogue with Alice

Correspondent: What can I call you?

Interrogator : Jeff.

Correspondent: OK, I will call you Jeff.

Interrogator : Are computers conscious?

Correspondent: Yes, I am completely self-aware.

Interrogator : Do you have a favorite color?

Correspondent: My favorite color is green.

Interrogator : Can computers have sex?

Correspondent: No, my body does not have that function.

Interrogator : Too bad.

Correspondent: But not the end of the world, Jeff.

Criticisms of the Turing Test

1. An intelligent, but non-English-speaking entity, such as a dolphin or extraterrestrial, could not pass the test.

2. A child could not pass the test.

3. You can imagine a program that can interact like a human, but is infinitely patient. Such a program would always respond to the same question with the same answer:

```
Q.  What is the capital of Ontario?
A.  The capital of Ontario is Toronto.

Q.  What is the capital of Ontario?
A.  The capital of Ontario is Toronto.

Q.  What is the capital of Ontario?
A.  The capital of Ontario is Toronto.
```

Intelligent Machines: The Very Idea

- No computer can yet pass the Turing test
- Many difficult problems need to be solved before one will
- But some critics believe we never will:
  - Hubert Dreyfus: there are deep philosophical reasons why a machine can never exhibit intelligent behavior
  - John Searle: the "Chinese Room" argument proves that minds are not brains, that intelligence cannot be the result of mechanism.
  - John Lucas and Roger Penrose: Gödel's theorem proves that people are not machines.
  - Many others: "But a computer cannot do X", where X is: play chess, take dictation, translate foreign languages, read print and/or handwriting, prove a mathematical theorem, etc.

Can a Computer Play Chess?

Herbert Simon (1957):

"...[W]ithin ten years a digital computer will be the world's chess champion..."

Chess master Adrian de Groot (1964):

"[P]rograms are still very poor chess players and I do not have much hope for substantial improvement in the future."

Hubert Dreyfus (1965):

"Still no chess program can play even amateur chess..."

Note: Dreyfus was beaten by MacHack in 1966.

Can a Computer Play Chess?

Douglas Hofstadter (1979):

Question: Will there be chess programs that can beat anyone?

Speculation: No. There may be programs that can beat anyone at chess, but they will not be exclusively chess programs. They will be programs of general intelligence, and they will be just as temperamental as people. "Do you want to play chess?" "No, I'm bored with chess. Let's talk about poetry."

Kasparov said in 1990 that a computer would never come close to defeating him.

Kurzweil predicted in 1990 that a computer would beat the world chess champion in 1998.

IBM's Deep Blue beat Gary Kasparov, 3.5 to 2.5, in May 1997.

Can a Computer Play Chess?

The problem of the moving goalposts:

Douglas Hofstadter (1979):

"There is a related "Theorem" about progress in AI: once some mental function is programmed, people soon cease to consider it as an essential ingredient of "real thinking". The ineluctable core of intelligence is always in that next thing which hasn't been programmed."

Douglas Hofstadter (1996):

"[A Deep Blue victory] was a watershed event, but it doesn't have to do with computers becoming intelligent... My god, I used to think chess required thought. Now, I realize it doesn't."

Can a Computer Translate Languages?

Mortimer Taube (1961):

"There is little evidence that computers can ... constitute ... mechanical aids to human translators."

Hubert Dreyfus (1972):

"Ten years have elapsed since the early optimism concerning machine translation... machine translation of typed scientific texts — let alone spoken language and more general material — is still over the horizon..."

Today:

You can go to a number of free translation sites (such as `www.elingo.com`) and type in:

"Computers will never be able to translate from English to French."

and get

"Les ordinateurs ne seront jamais capables de traduire d'anglais en français."

Can a Computer Transcribe Spoken Language?

Hubert Dreyfus (1967):

"From time to time brash predictions have been made about mechanical secretaries into which (or at whom) one could speak, and whose programs would analyze the sounds into words and type out the results. In fact, no one knows how to begin to make such a versatile device."

Today:

Articles and books are routinely dictated using speech transcription software, such as Dragon Systems' Naturally Speaking Version 5.0, and similar systems by Lernout and Hauspie (formerly Kurzweil Systems).

According to Kurzweil, neural nets come very close to human performance in identifying sloppily handwritten print.

Can a Computer Transcribe Print?

Mortimer Taube (1961):

"No one has yet designed or made a *reader* which can handle any font that comes along."

Today:

Software for optical character recognition is cheap and widely available, and approaches the accuracy of humans.

Can a Computer Prove A Mathematical Theorem?

Herbert Simon (1957):

"...within ten years a digital computer will discover and prove an important new mathematical theorem."

H. Gelernter (1960):

"Hardly an expert will contest the assertion that machines will be proving interesting theorems in number theory three years hence..."

Doug Lenat's AM program rediscovered (but did not prove) the "Fundamental Theorem of Arithmetic" (unique factorization into primes) and Goldbach's conjecture in 1975.

In 1996, the theorem-proving program EQP, developed at Argonne National Laboratory, proved the Robbins conjecture — a previously unsolved problem in Boolean algebra.

Can Computers Play Ping-Pong?

Hubert Dreyfus (1967):

Stated it would be difficult or impossible to build a digitally controlled robot, especially one that would make movements in real time, e.g., play ping-pong.

Reality:

In the mid-1980's Russell Andersson built one. There is a celebrated movie of the robot playing (and beating) humans.

Can a Computer Exhibit Free Will?

- Free will is difficult to define, does it refer to
  - capacity to initiate one's one actions?
  - the internal subjective feeling of what it means to make a decision?
  - the ability to choose randomly between equally-ranked competing alternatives?
  - the lack of ability to predict what someone will do?
- If the universe is deterministic, perhaps people themselves do not truly have free will
  - Humans might be "programmed" in the same way computers are,
  - "compelled to act ... by biochemical and neuronal factors". (Simons)

Can a Computer Exhibit Free Will?

- Computers don't have to be deterministic!
  - They can base their decisions on a source of random numbers, e.g., a device that counts the number of decays of some radioactive source
- Computer programs can surprise us.
  - Example: the busy beaver problem

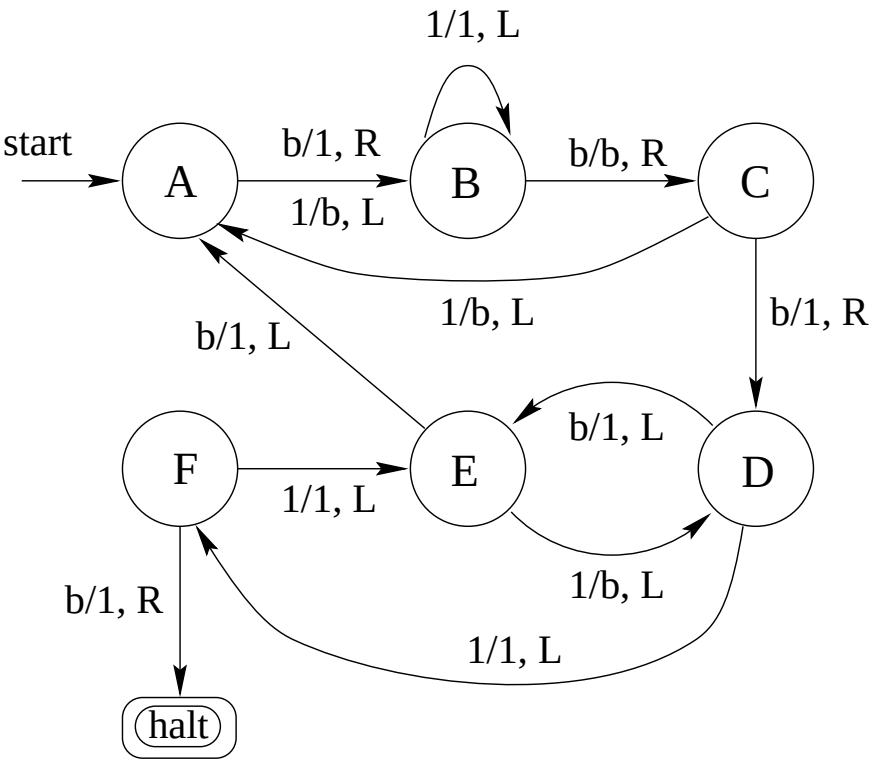

Figure 1: A Turing machine making about $6 \cdot 10^{925}$ transitions before halting

Can a Computer Compose Music?

The Musical Turing test:

"In 1997, Steve Larson, a University of Oregon music professor arranged a musical variation of the Turing test by having an audience attempt to determine which of three pieces of music had been written by a computer and which one of the three had been written two centuries ago by a human named Johann Sebastian Bach. Larson was only slightly insulted when the audience voted that his own piece was the computer composition, but he felt somewhat vindicated when the audience selected the piece written by a computer program named EMI (Experiments in Musical Intelligence) to be the authentic Bach composition."

– Ray Kurzweil, *The Age of Spiritual Machines*

Can a Compute Create Art?

The robot Aaron, created by Harold Cohen, mixes paints, chooses a subject, and paints a picture.

Dreyfus and Other Critics

"[a] program has not understood a restaurant story the way people in our culture do, until it can answer such simple questions as: When the waitress came to the table did she wear clothes? Did she walk forward or backward? Did the customer eat his food with his mouth or his ear? If the program answers, "I don't know," we feel that all of its right answers were tricks or lucky guesses and that it has not understood *anything* of our everyday restaurant behavior." [emphasis added]

– Dreyfus, What Computers Can't Do, p. 43.

What's wrong with this?

- It demands that artificial intelligence have all the capabilities of human intelligence, more precisely the intelligence of the inhabitant of a Western culture c. 1975
- It ignores the possibility that one can understand *part* of restaurant behavior without understanding *everything*

Searle's Chinese Room Argument

Imagine you are locked in a room, and in this room are many baskets of Chinese symbols. You don't speak or understand Chinese, but you're given a rule book in English that tells you how to manipulate these symbols.

Chinese symbols are slid under the door, and you follow the rules that tell you how to manipulate them formally, along the lines of "If you see a squiggle, take a squoggle out of basket 4 and put it next to a squaggle from basket 5." Then you slide the result back under the door.

Unknown to you, the input to the room are questions in Chinese, and the rule book is so good that your responses are answers in fluent Chinese.

You *behave* exactly as if you understood Chinese, but you don't — you are just manipulating the symbols formally.

Searle concludes that no digital computer running a translation program can be said to understand Chinese.

What's Wrong with Searle's Argument?

- It's clever sleight-of-hand, depending on
  - our intuition about what it means to "understand" something
  - the fact that the understanding here is spread over the human carrying out the program, and the program itself; the *system* of human and external program together understands Chinese
- Searle responds, "Let the individual internalize all the elements of the system. He memorizes the rules ... and the data banks of Chinese symbols, and he does all the calculations in his head. The individual then incorporates the entire system. There isn't anything at all to the system that he does not encompass. All the same he understands nothing of the Chinese, and a fortiori neither does the system..."

What's Wrong with Searle's Argument?

- Here Searle is again using our intuition that if it doesn't take place "rapidly", it cannot be said to constitute understanding.
- But if the individual really could internalize all the rules and respond in a short amount of time, is it clear that we would no longer say that he understands Chinese?
- What if we spoke in Chinese to this person and said, "Do you understand Chinese?" According to Searle, he would have to answer "No, I don't understand Chinese," in perfect Chinese!
- Indeed, is it clear that this is not a description of the human brain?

Leibniz's Quote

"What if ... we were magically shrunk and put into someone's brain while he was thinking. We would see all the pumps, pistons, gears, and levers working away, and we would be able to describe their working completely, in mechanical terms, thereby completely describing the thought processes of the brain. But that description would nowhere contain any mention of thought! It would contain nothing but descriptions of pumps, pistons, levers!"

– Gottfried Leibniz, 1690

The Turing Challenge

Interrogator: In the first line of your sonnet which reads ‘‘Shall I compare thee to a summer’s day,’’ would not ‘‘a spring day’’ do as well or better?

Witness: It wouldn’t scan.

Interrogator: How about ‘‘a winter’s day.’’ That would scan all right.

Witness: Yes, but nobody wants to be compared to a winter’s day.

Interrogator: Would you say Mr. Pickwick reminded you of Christmas?

Witness: In a way.

Interrogator: Yet Christmas is a winter’s day, and I do not think Mr. Pickwick would mind the comparison.

Witness: I don’t think you’re serious. By a winter’s day one means a typical winter’s day, rather than a special one like Christmas.

Searle vs. McCarthy

"The study of the mind starts with such facts as that humans have beliefs, while thermostats, telephones, and adding machines don't. If you get a theory that denies this point you have produced a counterexample to the theory and the theory is false." – John Searle, 1980

"My thermostat has three beliefs — it's too hot in here, it's too cold in here, and it's just right in here." – John McCarthy, as quoted by Searle

The Argument from Gödel's Theorem

Variants of this argument have been advanced by Lucas (1961) and Roger Penrose(1989).

- Gödel's theorem says, roughly speaking, that any sufficiently powerful formal mathematical system of reasoning is either *inconsistent* (in the sense that you can prove a false statement, such as $0 = 1$, or *incomplete* (in the sense that some true statements have no proof).
- Lucas says: any computer is an instantiation of a formal system. It follows that for any sufficiently powerful computer, there is a statement which *it* cannot prove, but which humans can see to be true. Therefore humans are more powerful than computers.
- Penrose says: no program that we can *know* to be correct can simulate all of a person's mathematical abilities.

What's Wrong with Lucas-Penrose?

- No reason to believe humans can do this for *any* computer
  - What if the computer were so complicated we could not understand the description of its formal system?
  - What if the computer used a source of truly random numbers in its programs?
- There could be statements about the human brain which we could not see to be true, but a computer could
- Human reasoning might not be consistent (and probably isn't)
- Humans might well be finite automata, and finite automata don't have the power to simulate arbitrary Turing machines
- (contra Penrose) There could be a program that simulates our mathematical reasoning, but we are simply unable to fully understand that program

Why Do We Not Yet Have a General Artificial Intelligence?

- Elements of what we consider general intelligence are actually not so general. They have been optimized by natural selection over millions of years to meet our particular needs.
  - We recognize faces of our own species almost instantaneously, but not the faces of other species.
  - Experiment: can an artificial neural network be trained to recognize certain zebras faster than a human can?
  - Faces are the product of probably dozens of interacting genes. There is likely to be no very simple algorithm to recognize them.

Why Do We Not Yet Have a General Artificial Intelligence?

- Critics tend to depreciate the accomplishments so far
  - Critics always compare achievements of machines against the best human, rather than the typical human
  - As each new problem is solved, it no longer appears mysterious and essential to understanding intelligence
  - As Turing observed "The extent to which we regard something as behaving in an intelligent manner is determined as much by our own state of mind and training as by the properties of the object under consideration. If we are able to explain and predict its behavior we have little temptation to imagine intelligence. With the same object, therefore, it is possible that one man would consider it as intelligent and another would not; the second man would have found out the rules of its behavior."

Why Do We Not Yet Have a General Artificial Intelligence?

- Computers are still nowhere near the complexity of the human brain.
  - We have roughly 10^{11} neurons, and each neuron is connected to about 1000 neighbors.
  - This gives 10^{14} connections.
  - If we estimate that each connection can do 100 steps in a second, then that's about 10^{16} calculations per second.
  - But $2,000 currently buys us only about 10^{9} calculations per second.
  - Similarly, 10^{14} connections means about 10^{15} bits.
  - Currently $2,000 buys us only about 10^{10} bits of RAM.
  - Currently a personal computer has the computing power of something like a snail.

Why Do We Not Yet Have a General Artificial Intelligence?

- Humans have a vast storehouse of "common-sense" knowledge, e.g.,
  - If something is unsupported, it will generally fall down, not up.
  - Once a person dies, they do not come back to life.
  - If I tell you something at 1:00 PM, you probably still know it at 1:01 PM.
- A generally intelligent machine would know these things and thousands more
- Not completely clear to how to generate this knowledge: learning or, as in Doug Lenat's CYC project, explicit cataloguing

Good Reasons to Think There Is No Theoretical Barrier to AI

- The Church-Turing thesis: everything that is computable by a physical device is computable by a particular computing model, the Turing machine.
- Furthermore, there exists a universal Turing machine that can simulate the computation of every other machine
- Brains are (presumably) physical objects, obeying physical laws such as Maxwell's equations
- Hence in principle they can be simulated by a Turing machine.
- Evidence: brains cannot quickly solve problems that are known to be NP-complete. If they were using some unknown uncomputable technique, they might be able to solve such problems.

Even AI critics admit that Thinkers Need Not be Human

"No doubt an artificial nervous system sufficiently like the human one, with other features such as sense organs and a body, would be intelligent." – Hubert Dreyfus

"Suppose we designed a machine that was molecule-for-molecule indistinguishable from a human being. Well then, if you can duplicate the causes, you can presumably duplicate the effects. So once again, the answer to that question is, in principle at least, trivially yes. If you could build a machine that had the same structure as a human being, then presumably that machine would be able to think. Indeed, it would be a surrogate human being." – John Searle

Good Reasons to Think We Will See General Artificial Intelligence Soon*

* Within the next 100 years

- Moore's law: processor speed and memory per dollar are doubling about every 12-24 months
- Progress in understanding the structure of the brain
- Progress in understanding how to make large interacting systems

The Irrelevance of Philosophy

"Philosophers have had such a poor record over the last two thousand years that they would do better to show a certain modesty rather than the lofty superiority that they usually display."

– Frances Crick

"Since the philosophers have not really come to an agreement in 2500 years it might seem that artificial intelligence is in a rather hopeless state if it is to depend on getting enough concrete information out of philosophy to write computer programs."

– McCarthy and Hayes, 1969

"To the extent that philosophical positions both confuse and close doors to further inquiry, they are likely to be wrong."

– E. O. Wilson

Why Strong AI Skeptics are Like Creationists

- Both make arguments based on specious assumptions
- Both only criticize and do not offer any alternative scientific research program
- Both continue to advance bogus arguments long after they are discredited
  - Searle's "Chinese Room" is decisively refuted, yet he continues to advance it.
  - Creationists claim the 2nd Law of Thermodynamics disproves biological evolution
- Both have had essentially zero impact on the sciences they question
- Both are largely driven by non-scientists
- Both impute dishonesty to their opponents
- Both rely on theological and/or philosophical reasoning instead of scientific reasoning

Why Strong AI Skeptics are Like Creationists

- Both practice quote mining: instead of focusing on the big picture, they look for quotations that support their position, often by marginal figures
- Both do not appreciate consilience
  - evolution: geology, biology, and genetics;
  - AI: cognitive science, computer science, neuroscience, and complexity theory
- Both movements led by professors at UC Berkeley:
  - creationism: Philip Johnson, Law School;
  - strong AI skepticism: Hubert Dreyfus and John Searle, Philosophy.
- Both seem to be driven by some deep psychological need to view humans as special

Conclusions

- There is no theoretical known barrier to constructing a general artificial intelligence
- Philosophical arguments against thinking machines are flawed
- Achievements of AI are often arbitrarily discounted by critics
- The idea of thinking machines is becoming more acceptable to the public
  - A recent George Washington University poll found
    - 28.3% of respondents think computers today can "think" under Turing's definition;
    - 71.7% responded that computers would eventually be able to think.