# The Turing Test: Then and Now

Peter Hawke

Stanford Philosophy
phawke@stanford.edu

October 16, 2012

# Some references


1. Shieber, ed. *The Turing Test: Verbal Behavior as the Hallmark of Intelligence*, MIT press, (2004).
2. Moor, ed., *The Turing Test: The Elusive Standard of Artificial Intelligence*, Kluwer Academic Publishers (2003).
3. Epstein et al. eds., *Parsing the Turing Test*, Springer (2008).
4. Abramson, Darren, "Turing's Responses to Two Objections", in *Minds and Machines* 18:147-267 (2008).

# Menu

What we'll talk about today, time permitting:

1. Outline of the Turing Test.
2. **History:** precursors to the Turing Test.
3. **History:** A few points of scholarly controversy surrounding Turing's famous 1950 paper.
4. **Recent history:** the Loebner competition.
5. Some **philosophical issues** raised by the Turing Test.
6. Some **classic objections** to the Test: methodological objections; the Lovelace objection; the Chinese Room; the Mathematical objection.
7. Consideration of **recent work** on these objections - showing that these issues are still live, and there is certainly scope for further discovery and understanding.

# The Turing Test, Rough Version

What is the Turing Test?

# The Turing Test, Rough Version

What is the Turing Test?

*The rough idea:* a good way to test if a computer is genuinely intelligent is to check if in its verbal behavior it is virtually indistinguishable from a human being.

**"Verbal Behavior as the Hallmark of Intelligence".** (title of Shieber book)

The use of 'verbal' here is slightly archaic: behavior is verbal if it is spoken or written behavior.

# The Turing Test, More Precise Version

## What is the Turing Test?

More precisely: there is a contestant and a judge. The contestant is either a computer or a human being. The judge and the contestant cannot observe each other, and can only communicate via written word eg. by sending natural language messages to each other, via a computer prompt. The judge can communicate with the contestant about any topic they like, and ask questions and make claims in any manner they like. The contestant passes the test if the judge, after communicating with the contestant for a fair amount of time, is happy to predict that the contestant is a human being. The contestant fails if they are predicted to be a computer.

# The Turing Test, More Precise Version

## What is the Turing Test?

More precisely: there is a contestant and a judge. The contestant is either a computer or a human being. The judge and the contestant cannot observe each other, and can only communicate via written word eg. by sending natural language messages to each other, via a computer prompt. The judge can communicate with the contestant about any topic they like, and ask questions and make claims in any manner they like. The contestant passes the test if the judge, after communicating with the contestant for a fair amount of time, is happy to predict that the contestant is a human being. The contestant fails if they are predicted to be a computer.

Practical decisions to make: who should be the judge? Would it be better to have more than one judge? How long should the conversation last?

# The Turing Test, More Precise Version

What is the Turing Test?

*A key claim: if a computer contestant is consistently ruled to be human in the TT, then we may conclude that the computer is as intelligent as a human being.*

# The Turing Test, More Precise Version

What is the Turing Test?

*A key claim: if a computer contestant is consistently ruled to be human in the TT, then we may conclude that the computer is as intelligent as a human being.*

Practical question: how well exactly does the computer need to perform to be declared intelligent? Does it need to pass the test the same percentage of time that a human being passes it? Should we take that to be "100 percent of the time"?

# Some history

Let's begin our historical remarks with the inventor of the Turing Test.

# Some history

Let's begin our historical remarks with the inventor of the Turing Test.

No, it's not Alan Turing (not exactly, anyway).

# From Descartes' *Discourse on the Method*, 1637:

*If there were machines which bore a resemblance to our bodies and imitated our actions as closely as possible for all practical purposes, we should still have two very certain means of recognizing that they were not real men. The first is that they could never use words, or put together signs, as we do in order to declare our thoughts to others. For we can certainly conceive of a machine so constructed that it utters words, and even utters words that correspond to bodily actions causing a change in its organs. But it is not conceivable that such a machine should produce different arrangements of words so as to give an appropriately meaningful answer to whatever is said in its presence, as the dullest of men can do. Secondly, even though some machines might do some things as well as we do them, or perhaps even better, they would inevitably fail in others, which would reveal that they are acting not from understanding, but only from the disposition of their organs. For whereas reason is a universal instrument, which can be used in all kinds of situations, these organs need some particular action...* (Translation by Robert Stoothoff)

# Some history

- An important precursor to Turing was Descartes.
- Clockwork machinery was taking off in Descartes' time.
- This inspired the *design argument* in philosophy of religion, for one thing.
- Descartes was interested in the distinction between humans (souls) and animals (no souls).
- He subscribed to the doctrine of the *bête machine* (beast machine): animals are mere machines.
- So he needed a clear criterion to separate humans from machines.
- Some of his Cartesian followers, like De Cordemoy, further explored this issue.

*Source: Copeland 2000, Shieber 2004*

# Surprising facts?

It is generally taken that Turing introduces the Turing Test, properly speaking, and the associated criterion for intelligence, in his classic 1950 paper "Computing Machinery and Intelligence", originally published in *Mind*.

# Surprising facts?

It is generally taken that Turing introduces the Turing Test, properly speaking, and the associated criterion for intelligence, in his classic 1950 paper "Computing Machinery and Intelligence", originally published in *Mind*.

But surprising facts (because they are easy to forget or overlook):

1. Turing does not call the test he talks about in this paper the "Turing Test".
2. The test he describes in this work appears to be *significantly different* to the test just described.

# The name

A potted early history of AI:

- Around 1941: Turing circulates a type-written paper on machine intelligence.
- 1950: "Computing Machinery and Intelligence"
- 1952: Earliest AI programs written and run in England.
- 1956: The Logic Theorist by Newell, Simon and Shaw.
- 1956: McCarthy's *Dartmouth Summer Research Project on Artificial Intelligence*
- 1960's: various articles in *Mind* discussing Turing's Imitation Game.
- 1976: James Moor writes a classic paper defending the "Turing Test".

*Source: Copeland 2000, Saygin et al. 2000*

# The Imitation Game

Here is the test that Turing outlines in his 1950 paper:

# The Imitation Game

Here is the test that Turing outlines in his 1950 paper:

The Imitation Game: Consider the following game. There are three participants: a man or computer (A), a woman (B) and an interrogator (C). They can communicate only by typed written word, and cannot otherwise observe each other. The objective of the interrogator is to determine which of A and B is the woman. The aim of A is to win, so must do his best to imitate a woman. C may ask any questions s/he likes, including direct questions like "are you a woman?" (Of course, A need not answer truthfully).

# The Imitation Game

Here is the test that Turing outlines in his 1950 paper:

The Imitation Game: Consider the following game. There are three participants: a man or computer (A), a woman (B) and an interrogator (C). They can communicate only by typed written word, and cannot otherwise observe each other. The objective of the interrogator is to determine which of A and B is the woman. The aim of A is to win, so must do his best to imitate a woman. C may ask any questions s/he likes, including direct questions like "are you a woman?" (Of course, A need not answer truthfully).

The criterion for intelligence: if a computer can perform in this game as well as a man can, then that computer is intelligent.

# Turing Test or Imitation Game?

- It is somewhat ambiguous whether Turing has the *original imitation game* (OIG) or the Turing Test (TT) in mind in his 1950 paper. This is a matter of scholarly dispute.
- Later, in a 1952 radio broadcast, he unequivocally describes the TT as the test he has in mind. Copeland 2000 thinks this settles the scholarly dispute.

# Turing Test or Imitation Game?

But does it matter?

Sterrett 2000 thinks that it does. The two tests are not equivalent in their characteristics. She argues that the OIG in fact has some nice features that the TT lacks.

# Turing Test or Imitation Game?

But does it matter?

Sterrett 2000 thinks that it does. The two tests are not equivalent in their characteristics. She argues that the OIG in fact has some nice features that the TT lacks.

1. "...[I]ntelligence lies, not in having the cognitive habits developed in learning to converse, but in the exercise of the intellectual powers required to recognize, evaluate, and, when called for, override them".

# Turing Test or Imitation Game?

But does it matter?

Sterrett 2000 thinks that it does. The two tests are not equivalent in their characteristics. She argues that the OIG in fact has some nice features that the TT lacks.

1. "... [I]ntelligence lies, not in having the cognitive habits developed in learning to converse, but in the exercise of the intellectual powers required to recognize, evaluate, and, when called for, override them".
2. The OIG allows for a control group (of men) such that we can check to see if a computer can do *better* than the control group.

# Turing Test or Imitation Game?

But does it matter?

Sterrett 2000 thinks that it does. The two tests are not equivalent in their characteristics. She argues that the OIG in fact has some nice features that the TT lacks.

1. "...[I]ntelligence lies, not in having the cognitive habits developed in learning to converse, but in the exercise of the intellectual powers required to recognize, evaluate, and, when called for, override them".
2. The OIG allows for a control group (of men) such that we can check to see if a computer can do *better* than the control group.
3. There is evidence that a judge in an actual Turing Test tends to be biased against the machine. "In the Loebner competition held at Dartmouth College in January 2000, human respondents were mistaken for computers on 10 occasions, a computer for a human on none." (Copeland, 2000).

# The Loebner Competition

Recent history: Turing Tests are regularly carried out nowadays, at annual Loebner Contest, since 1991.

# The Loebner Competition

Recent history: Turing Tests are regularly carried out nowadays, at annual Loebner Contest, since 1991.

Notable: the machines that compete do not have a very impressive record for fooling the judges.

Moor 2000: "[At the 2000 Dartmouth Loebner contest:] No computer was mistaken for a human though in a few cases humans were judged to be computers!"

# The Loebner Competition

Recent history: Turing Tests are regularly carried out nowadays, at annual Loebner Contest, since 1991.

Notable: the machines that compete do not have a very impressive record for fooling the judges.

Moor 2000: "[At the 2000 Dartmouth Loebner contest:] No computer was mistaken for a human though in a few cases humans were judged to be computers!"

Turing, 1950: "I believe that in about fifty year' time it will be possible to programme computers . . . to make them play the imitation game so well than average interrogator will not have more than 70 percent chance of making the right identification after five minutes of questioning."

The 2000 Loebner competition is taken as evidence that this prediction has not panned out.

# The Loebner Contest

A compelling conversation:

**Judge:** where do you work?
**Program:** At the university. I mostly clean the Standish building.
**Judge:** What university?
**Program:** University of Eastern Ontario. I'm a cleaner. I empty trash. Vacuum. Dust. You know.
**Judge:** And why do they want to fire you?
**Program:** It's just because my boss, Mr. Johnson, is mad at me.
**Judge:** Why is he mad at you?
**Program:** I didn't do nothing wrong. But he thinks that I did. It's all because of that radio they think I stole from the lab.
**Judge:** That's too bad. Are you in a union?
**Program:** What would I tell the union? They won't say they're firing me because I stole the radio. . .

A not so compelling conversation:

**Judge:** What is Einstein's first name?

**Program:** Sometimes your questions are very strange. Is it a kind of puzzle?

... **Program:** ...do you think you could tell me your name now?

**Judge:** Alexander

**Program:** Nice to meet you, Alexander

**Judge:** Do you like the arts?

**Program:** I'm not entirely sure what you mean. Anyway...I assume you're male, Alexander - am I right?

# Connected philosophical issues

Here are some important philosophical issues raised by Turing's proposal of the Turing Test:

# Connected philosophical issues

Here are some important philosophical issues raised by Turing's proposal of the Turing Test:

1. Is building a machine that can pass the TT a sensible goal for AI to have (putting aside if it is a good test for intelligence and whether machines can be intelligent)?
   - Is the test too narrow?
   - Is the test too hard?

# Connected philosophical issues

Here are some important philosophical issues raised by Turing's proposal of the Turing Test:

1. Is building a machine that can pass the TT a sensible goal for AI to have (putting aside if it is a good test for intelligence and whether machines can be intelligent)?
   - Is the test too narrow?
   - Is the test too hard?
2. Is the TT a good test for intelligence (putting aside the issue as to whether it is possible for computers to be intelligent)?
   - Is passing the test logically sufficient for being intelligent?
   - Is passing the test good inductive evidence for intelligence?

# Connected philosophical issues

Here are some important philosophical issues raised by Turing's proposal of the Turing Test:

1. Is building a machine that can pass the TT a sensible goal for AI to have (putting aside if it is a good test for intelligence and whether machines can be intelligent)?
   - Is the test too narrow?
   - Is the test too hard?
2. Is the TT a good test for intelligence (putting aside the issue as to whether it is possible for computers to be intelligent)?
   - Is passing the test logically sufficient for being intelligent?
   - Is passing the test good inductive evidence for intelligence?
3. Is it possible for a computer to be intelligent?

# Connected philosophical issues

Here are some important philosophical issues raised by Turing's proposal of the Turing Test:

1. Is building a machine that can pass the TT a sensible goal for AI to have (putting aside if it is a good test for intelligence and whether machines can be intelligent)?
   **Objection:** Ford and Hayes' methodological objections.

# Connected philosophical issues

Here are some important philosophical issues raised by Turing's proposal of the Turing Test:

1. Is building a machine that can pass the TT a sensible goal for AI to have (putting aside if it is a good test for intelligence and whether machines can be intelligent)?
   **Objection:** Ford and Hayes' methodological objections.
2. Is the TT a good test for intelligence (putting aside the issue as to whether it is possible for computers to be intelligent)?
   **Objection:** The Lovelace Objection.

# Connected philosophical issues

Here are some important philosophical issues raised by Turing's proposal of the Turing Test:

1. Is building a machine that can pass the TT a sensible goal for AI to have (putting aside if it is a good test for intelligence and whether machines can be intelligent)?
   **Objection:** Ford and Hayes' methodological objections.
2. Is the TT a good test for intelligence (putting aside the issue as to whether it is possible for computers to be intelligent)?
   **Objection:** The Lovelace Objection.
3. Is it actually possible for a computer to be intelligent or think?
   **Objection:** Searle's Chinese Room; the Mathematical Objection.

# Connected philosophical issues

Here are some important philosophical issues raised by Turing's proposal of the Turing Test:

1. Is building a machine that can pass the TT a sensible goal for AI to have (putting aside if it is a good test for intelligence and whether machines can be intelligent)?
   **Objection:** Ford and Hayes' methodological objections.
2. Is the TT a good test for intelligence (putting aside the issue as to whether it is possible for computers to be intelligent)?
   **Objection:** The Lovelace Objection.
3. Is it actually possible for a computer to be intelligent or think?
   **Objection:** Searle's Chinese Room; the Mathematical Objection.

# Methodological Objections

"The Turing Test had a historical role in getting AI started, but it is now a burden to the field, damaging its public reputation and its own intellectual coherence. We must explicitly reject the Turing Test in order to find a more mature description of our goals; it is time to move it from the textbooks to the history books" (Hayes and Ford, 1995)

# Methodological Objections

Some of their more particular concerns are:

# Methodological Objections

Some of their more particular concerns are:

1. The TT at best tests for *human* intelligence. But should the ultimate goal of AI really be to simply recreate what a human does well and what a human does badly?

# Methodological Objections

Some of their more particular concerns are:

1. The TT at best tests for *human* intelligence. But should the ultimate goal of AI really be to simply recreate what a human does well and what a human does badly?
2. The TT is of no use in judging AI's success in areas like vision and robotics.

# Methodological Objections

Some of their more particular concerns are:

1. The TT at best tests for *human* intelligence. But should the ultimate goal of AI really be to simply recreate what a human does well and what a human does badly?
2. The TT is of no use in judging AI's success in areas like vision and robotics.
3. Focus on passing the TT encourages programming tricks. Instead, we should focus on the fundamental principles of reasoning.

# Methodological Objections

Some of their more particular concerns are:

1. The TT at best tests for *human* intelligence. But should the ultimate goal of AI really be to simply recreate what a human does well and what a human does badly?
2. The TT is of no use in judging AI's success in areas like vision and robotics.
3. Focus on passing the TT encourages programming tricks. Instead, we should focus on the fundamental principles of reasoning.

# Methodological Objections

To illustrate the last point, every program in the 2000 Loebner competition answered the following questions wrong (while the human contestants got them all right) (Moor 2000):

1. What is the color of a blue truck?
2. Where is Sue's nose when Sue is in her house?
3. What happens to an ice cube in a hot drink?
4. Altogether how many feet do four cats have?
5. How is the father of Andy's mother related to Andy?
6. What letter does the letter 'M' look like when turned upside down?
7. What comes next after A1, B2, C3?
8. Reverse the digits in 41.
9. PLEASE IMITATE MY TYPING STYLE.

# Methodological Objections

Moor's reply (Moor 2000):

No one, including Turing, claims that AI can't have diverse goals and tests.

Moor's reply (Moor 2000):

*Intelligence Attribution Argument:* The TT is at least a useful thought experiment, intended to establish that machine intelligence is a *coherent possibility*. Consider this argument: suppose that a machine passes the TT. In the case of human beings, we take this level of verbal behavior to be excellent inductive evidence that other human beings have minds (rejecting this gives us solipsism). By parity of reasoning, there is excellent inductive evidence that our machine has a mind. Hence, if we wish to avoid solipsism, then we must accept that attributing intelligence to a machine is a coherent possibility.

# Methodological Objections

Moor's reply (Moor 2000):

*The Methodology Argument:* The details of the TT are less important than the idea: focus on *behavior* to establish if a machine has the "mental characteristics" we wish it to have.

# Methodological Objections

Moor's reply (Moor 2000):

*The Visionary Argument:* The TT is an inspiring ideal goal for AI.

# Is the TT a good test for intelligence?

Turing addresses the Lovelace objection in his paper. Attributed originally to Lady Lovelace (who had Babbage's analytical engine in mind), the objection goes:

> *Computers can't create anything. For creation requires, minimally, originating something. But computers originate nothing; they merely do that which we order them, via programs, to do. But if computers can't create/originate, then they can't be intelligent, thinking beings.*

# Is the TT a good test for intelligence?

Turing's reply:

1. Machines can "originate" in the sense of surprising their creators. Computers surprise their creators all the time.
2. A learning machine (eg. based on connectionist principles) can surprise its creator.

# Is the TT a good test for intelligence?

Turing's reply:

1. Machines can "originate" in the sense of surprising their creators. Computers surprise their creators all the time.
2. A learning machine (eg. based on connectionist principles) can surprise its creator.

The first objection looks like it's missing the point. A computer can surprise a programmer if the programmer made a mistake in programming, or if there is a random hardware fluke, or if the programmer simply doesn't understand the workings of the machine properly. But none of these indicate creativity.

# Is the TT a good test for intelligence?

Turing's reply:

1. Machines can "originate" in the sense of surprising their creators. Computers surprise their creators all the time.
2. A learning machine (eg. based on connectionist principles) can surprise its creator.

The first objection looks like it's missing the point. A computer can surprise a programmer if the programmer made a mistake in programming, or if there is a random hardware fluke, or if the programmer simply doesn't understand the workings of the machine properly. But none of these indicate creativity.
The second is better. But a problem remains: *one can imagine computers that aren't creative passing the TT*.

# Is the TT a good test for intelligence?

Bringsjord et al. 2001 suggests that the Lovelace Objection is best cashed out as suggesting a more satisfactory test for (machine) intelligence: **The Lovelace Test**

Artificial agent $A$, designed by $H$, passes LT if and only if

1. $A$ outputs $o$;
2. $A$'s outputting $o$ is not the result of fluke hardware error, but rather the result of processes that can repeat;
3. $H$ (or someone with her resources and knowledge) cannot explain how $A$ produced $o$ by appeal to $A$'s architecture, knowledge-base and core functions.

# Is the TT a good test for intelligence?

Notes:

- The Lovelace Test can be thought of in this way: for a machine to be intelligent is for it to hold a conversation with its creator that is not utterly predictable to its creator.

# Is the TT a good test for intelligence?

Notes:

- The Lovelace Test can be thought of in this way: for a machine to be intelligent is for it to hold a conversation with its creator that is not utterly predictable to its creator.
- Actually, Bringsjord et al. think that we can use the Lovelace Test to *prove* that no computer can be considered intelligent. Sketch: the Representation Theorem for first-order logic says that for any computable function there is a set of axioms $\Omega$ such that every value of that function can be derived from those axioms. But $\Omega$ is part of the knowledge-base. So in principle the creator of a machine can always predict what the machine will do.

# The Chinese Room

Is it possible for a machine to think? Searle 1980 denies this.

1. The Chinese Room is a representative computer.
2. It is obvious that the Chinese Room has no understanding of the strings of symbols it spews out.
3. There is no thinking or genuine intelligence without understanding.
4. Therefore: no computer can be intelligent or think.

Rapaport 2000 thinks that the important argument that emerges is this:

*The Semantic Argument*

1. Computer programs are purely syntactic.
2. Cognition is semantic.
3. Syntax alone is not sufficient for semantics.
4. Therefore: no purely syntactic computer program can exhibit semantic cognition.

Rapaport 2000 thinks that the important argument that emerges is this:

*The Semantic Argument*

1. Computer programs are purely syntactic.
2. Cognition is semantic.
3. Syntax alone is not sufficient for semantics.
4. Therefore: no purely syntactic computer program can exhibit semantic cognition.

One way to object is to deny the first premise, by arguing for the possibility of computers with appropriate relations to the external world. Rapaport denies the *third* premise, however.

# The Chinese Room

Rapaport:

- Syntax: study of relations between symbols in a system of symbols that the agent can cognitively manipulate.

# The Chinese Room

Rapaport:

- Syntax: study of relations between symbols in a system of symbols that the agent can cognitively manipulate.
- Semantics: study of relation between symbols and domain of meanings (objects not in the symbol system).

# The Chinese Room

Rapaport:

- Syntax: study of relations between symbols in a system of symbols that the agent can cognitively manipulate.
- Semantics: study of relation between symbols and domain of meanings (objects not in the symbol system).
- Claim: meanings for a thinking agent are themselves internal representations.

# The Chinese Room

Rapaport:

- Syntax: study of relations between symbols in a system of symbols that the agent can cognitively manipulate.
- Semantics: study of relation between symbols and domain of meanings (objects not in the symbol system).
- Claim: meanings for a thinking agent are themselves internal representations.
    1. Reason: we should accept representational realism. We don't have direct access to the world. All our cognitive activity - visual, conceptual, linguistic - deals in representations (eg. patterns of activated neural nodes).

# The Chinese Room

Rapaport:

- Syntax: study of relations between symbols in a system of symbols that the agent can cognitively manipulate.
- Semantics: study of relation between symbols and domain of meanings (objects not in the symbol system).
- Claim: meanings for a thinking agent are themselves internal representations.
  1. Reason: we should accept representational realism. We don't have direct access to the world. All our cognitive activity - visual, conceptual, linguistic - deals in representations (eg. patterns of activated neural nodes).
  2. Also: understanding a symbol involves relating it to a meaning (another representation) that is already understood. The chain of understanding must end with a domain of meanings that are understood in terms of themselves, otherwise either regress or no understanding.

# The Chinese Room

Rapaport:

- Syntax: study of relations between symbols in a system of symbols that the agent can cognitively manipulate.
- Semantics: study of relation between symbols and domain of meanings (objects not in the symbol system).
- Claim: meanings for a thinking agent are themselves internal representations.
  1. Reason: we should accept representational realism. We don't have direct access to the world. All our cognitive activity - visual, conceptual, linguistic - deals in representations (eg. patterns of activated neural nodes).
  2. Also: understanding a symbol involves relating it to a meaning (another representation) that is already understood. The chain of understanding must end with a domain of meanings that are understood in terms of themselves, otherwise either regress or no understanding.
- But that means we can always take union of "syntactic domain" and "domain of meaning", and consider this as pure syntax.

# The Mathematical Objection

Turing addresses the Mathematical Objection in his 1950 paper. The objection rests on a purported important difference between humans and machine due to the limitative results of logic and theoretical computer science, such as Goedel's incompleteness theorems and the Halting Problem. The idea of the objection is that machines have restrictions that humans don't and that it is plausible that humans escape these restrictions because of *intelligence*.

# The Mathematical Objection

Turing addresses the Mathematical Objection in his 1950 paper. The objection rests on a purported important difference between humans and machine due to the limitative results of logic and theoretical computer science, such as Goedel's incompleteness theorems and the Halting Problem. The idea of the objection is that machines have restrictions that humans don't and that it is plausible that humans escape these restrictions because of *intelligence*.

But what exactly *is* this objection? Abramson 2008 argues that Turing deals with *two distinct* such objections in his writing. And that scholars such as Piccinni, Copeland and Penrose mistakenly think he is dealing with the first, when he is dealing with the second in the 1950 paper.

# The Mathematical Objection

Turing 1947 'Lecture on the Automatic Computing Engine":

"It might be argued that there is a fundamental contradiction in the idea of a machine with intelligence. . . It has for instance been shown that with certain logical systems there can be no machine which will distinguish provable formulae of the system from unprovable, ie. that there is no test that the machine can apply which will divide propositions with certainty into these two classes. Thus if a machine made for this purpose *it must in some cases fail to give an answer*."

# The Mathematical Objection

But it seems that humans lack this restriction:

"On the other hand, if a mathematician is confronted with such a problem he would search around and find new method of proof, so that he ought eventually to be able to reach a decision about any given formula. This would be the argument."

# The Mathematical Objection

Turing's reply:

"Against it I would say that fair play must be given to the machine. Instead of it sometimes giving no answer we could arrange that it gives occasional wrong answers. But the human mathematician would likewise make blunders when trying out new techniques... In other words then, if a machine is expected to be infallible, it cannot also be intelligent."

So, the solution is to make computers fallible. Instead of always either giving a correct answer or not halting, the machine can be equipped to approximate answers using heuristics, sometimes on pain of being wrong. But this is exactly as an intelligent human behaves.

## The Mathematical Objection

Roger Penrose's version of this first mathematical argument from *Shadows of the mind*:

1. Let $C$ be a computer which, when provided with input $n$, either prints out a correct proof that $T_n(n)$ (the $n$th computer with input $n$) does not halt, or fails to halt. **(Infallibility)**.
2. Let $c$ be the index of $C$ in the enumeration of computers.
3. Assume $C(c)$ halts. Then it a produces a correct proof that $C(c)$ does not halt. This is contradictory, since we assumed that $C(c)$ halts.
4. Therefore, $C(c)$ does not halt.
5. Therefore, any intelligent person can see that $C(c)$ cannot halt (just follow the above argument) but $C$ is unable to prove this.

Turing's reply is: 1 will not be true for an intelligent computer, which has to be fallible. But in this case the reasoning in 3 does not go through.

# The Mathematical Objection

But is this first mathematical objection what Turing addresses in the 1950 paper? Abramson thinks not.

"There are a number of results of mathematical logic which can be used to show that there are limitations to the powers of discrete-state machines . . . The result in question [the Halting Problem] refers to a type of machine which is essentially a digital computer with an infinite capacity. It states that there are certain things that such a computer cannot do. If it is rigged up to give answers in the imitation game, **there will be some questions to which it will either give a wrong answer, or fail to give an answer at all however much time is allowed for a reply.**" (Turing 1950)

# The Mathematical Objection

Abramson 2008: Turing has outlined a different objection to the first mathematical objection.

# The Mathematical Objection

Abramson 2008: Turing has outlined a different objection to the first mathematical objection.

It is not a premise in the argument that the machine must be infallible. It is allowed that the computer can be fallible.

Turing concludes that, for any given machine, there is a question that can be posed to that machine such that the machine is guaranteed to give a *wrong answer*. The question is "will machine *n* answer yes to any question?', where *n* depends on the machine we are testing. Compare this to the other argument: the conclusion is that the computer will be forced, in some cases, to give *no answer* to a query.

# Conclusion

To sum up:

- The TT continues to generate live issues.
- Work is being done on issues directly related to the TT, and Turing's presentation of it.
- There are clearly interesting avenues for further research.

# The end

Thanks!