



---

**Judging Chatbots at Turing Test 2014**  
**Comments relating to the "Turing Test" event at the Royal Society**  
**in London UK, on 6-7th June 2014, by one of the "judges".**

**(DRAFT: Liable to change)**

**Aaron Sloman**  
**School of Computer Science, University of Birmingham.**  
**(Philosopher in a Computer Science department)**

---

This is one of two documents reflecting on the (mythical) Turing Test. The other is concerned with what can, in principle, be discovered about the capabilities of a complex information processing system by performing behavioural tests:

<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/black-box-tests.html>

---

[Jump to Table of Contents.](#)

---

**ABSTRACT**

When working on a general way to mechanise computation in 1936, Alan Turing did not propose a test for whether a machine has the ability to do computations. That would have been a silly thing to do, since no one test, e.g. with a fixed set of problems could be general enough. Neither did he propose what might be called a "test-schema" or "meta-test", namely calling in average people to give the machine computational problems and letting them decide whether it had succeeded, as might be done in some sort of reviewing process for a commercial product to help people with computation.

Instead, he proposed a deep THEORY, starting with an analysis of the sorts of things a computational machine needs to be able to do, based on the sorts of things human computers already did, e.g. when doing numerical calculations, algebra, or logic. He came up with a surprisingly simple generic specification for a class of machines, now known as Turing Machines, and demonstrated not by building examples and testing them, but mathematically, that any Turing machine would meet the requirements he had worked out. Since the specification allowed a TM to have any number of rules for manipulation, it followed mathematically that there are infinitely many types of Turing Machine, with different capabilities. Surprisingly, it also turned out that a special subset of TMs, the UNIVERSAL TMs (UTMs) could each model *any other TM*. Wikipedia provides a useful overview:

[http://en.wikipedia.org/wiki/Universal\\_Turing\\_machine](http://en.wikipedia.org/wiki/Universal_Turing_machine)

Doing something similar to what Turing did, but for machines that have human-like, or some other sort of intelligence, rather than for machines that merely (!) compute, would require producing a general theory about possible varieties of intelligence and their implications. This is totally different from the futile task of trying to define tests for machines with intelligence (or human-like intelligence)

-- as Turing himself understood. In Sloman (mythical) evidence based on the contents of Turing's 1950 paper is presented against the hypothesis that Turing intended to propose a test for intelligence: he was too intelligent to do any such thing. He was doing something much deeper.

I'll summarise the arguments below and describe an alternative research agenda, inspired by Turing's publications, including his 1952 paper, an agenda that he might have worked on had he lived longer, namely The Meta-Morphogenesis project, using evidence from multiple evolutionary trails to develop a general theory of forms of biological information processing, including many forms of intelligence.

Understanding the varieties of forms of information processing in organisms (including not only humans, but also microbes, plants, squirrels, crows, elephants, and orangutans) is a much deeper and more worthwhile scientific and philosophical endeavour than merely trying to characterise some arbitrary subset of that variety, as any specific behavioural test will do.

---

[Jump to Table of Contents.](#)

---

## **WORK IN PROGRESS**

**Installed:** 10 Jun 2014

**Last updated:**

29 Jun 2014; 4 Jul 2014; 25 Jul 2015 (Attempted to clarify a little.)

13 Jun 2014; 19 Jun 2014; 22 Jun 2014; 24 Jun 2014 (removed some repetition, added more headings);

This discussion note is

<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/turing-test-2014.html>

A slightly messy, automatically generated PDF version is:

<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/turing-test-2014.pdf>

(Or use your browser to generate one from the html file.)

Mary-Ann Russon interviewed me while it was being written and provided [a summary](#) in her International Business Times column.

For interest, and entertainment, here's a video of two instances of the same chatbot engaging in an unplanned theological discussion (at Cornell): <http://www.youtube.com/watch?v=WnzlbyTZsQY>

A partial index of discussion notes is in

<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/AREADME.html>

---

## **NOTE (Extended 19 Jun 2014)**

By the time you read this there will probably already be hundreds, or thousands, of web pages presenting comments on or criticisms of the Turing test in general, or the particular testing process at the Royal Society on 6-7 June 2014.

The announcement that one of the competing chatbots, Eugene Goostman, had been the first to pass the Turing Test produced an enormous furore. Here's a small sample of comments on the world wide web:

<http://www.reading.ac.uk/news-and-events/releases/PR583836.aspx>

<http://metro.co.uk/2014/06/09/actually-no-a-computer-did-not-pass-the-turing-test-4755769/>

<http://mashable.com/2014/06/12/eugene-goostman-turing-test/>

<http://www.bbc.co.uk/news/technology-27762088>

<http://iangent.blogspot.co.uk/2014/06/why-passing-of-turing-test-is-big-deal.html>

<http://www.ibtimes.co.uk/why-turing-test-not-adequate-way-calculate-artificial-intelligence-1452120>

[http://www.thestar.com/news/world/2014/06/13/was\\_the\\_turing\\_test\\_passed\\_not\\_everyone\\_thinks\\_so.html](http://www.thestar.com/news/world/2014/06/13/was_the_turing_test_passed_not_everyone_thinks_so.html)

<http://www.theguardian.com/commentisfree/2014/jun/11/ai-eugene-goostman-artificial-intelligence>  
<http://anilkseth.wordpress.com/2014/06/09/the-importance-of-being-eugene-what-not-passing-the-turing-test-really-means/>  
<http://www.wired.com/2014/06/beyond-the-turing-test/>  
<http://www.theverge.com/2014/6/11/5800440/ray-kurzweil-and-others-say-turing-test-not-passed>  
<http://www.newscientist.com/article/mg22229732.200-why-the-turing-test-needs-updating.html>  
<http://www.washingtonpost.com/news/morning-mix/wp/2014/06/09/a-computer-just-passed-the-turing-test-in-landmark-trial/>  
(Click here to see lots more!)

For some reason several reports in printed and online media referred to the winning contestant as a "supercomputer" rather than a program. Even if someone involved with the test produced that label in error, the inability of journalists to detect the error should be treated as evidence of poor journalistic education and resulting gullibility.

## **The idea of a behavioural test for intelligence is misguided, and Alan Turing did not propose one.**

I have previously criticised the idea of a 'turing test' for intelligence (in contrast with tests for a good theory about varieties of intelligence and how instances of that variety evolve and develop), and also criticised the suggestion that Turing was proposing such a test. Many who criticise the test try to offer improved versions of the test, e.g. requiring more than just textual interaction, or specifying a need for a lifelong test for an individual.

**All such proposals to improve on the test are misguided, for reasons explained below.**

I believe my main criticism of the idea of any sort of expanded behavioural test for intelligence below, based on a comparison with *trying to devise a test for computation*, is new, but would be grateful for links to web sites or documents that make the same points about the Turing Test as I do, and any that point out errors in this analysis.

Adam Ford (in Melbourne, Australia) interviewed me about this event using skype, on 12th June 2014 (2am UK time!) and has made the video available here (63 mins):

<http://www.youtube.com/watch?v=ACaJJcsvgL8>

It attempts to explain some of the points made below, in a question/answer session, but suffers from lack of preparation for the interview. Our interview at the AGI conference in Oxford in December 2012 is far superior: <https://www.youtube.com/watch?v=iuH8dC7Snn0>

---

## **CONTENTS**

- [ABSTRACT \(above\)](#)
- [NOTE \(Extended 19 Jun 2014 - above\)](#)
- [Why did I agree to be a judge in the London 2014 "Turing Test" event?](#)
- [Testing a generic, polymorphic design, not an instance.](#)
- [Criticisms and suggestions for improvement of the Turing Test](#)
- [Was the winning chatbot some sort of cheat?](#)
- [What did I learn about the state of the art?](#)
- [Educational value of chatbot development](#)
- [What's wrong with the Turing "Test"?](#)
- [Requirements for seeing](#)
- [Changing fashions of argumentation against the Turing test.](#)
- [The mythical Turing test.](#)
- [What were Turing's long term aims?](#)

- [More interesting and useful challenges than the \(mythical\) Turing Test?](#)
  - [Proposing a test for computation vs proposing a theory](#)
  - [Epigenetic conjectures](#)
  - **NOTE (CPHC RESPONSES):**
  - [References \(To be extended\)](#)
  - [THANKS](#)
- 

## Why did I agree to be a judge in the London 2014 "Turing Test" event?

For many years I have argued, like many AI researchers, that attempting to use the so-called Turing test to evaluate computer programs is a waste of time, and attempting to design machines to do well in public "Turing tests" does not significantly advance AI research, even though some of the results are mildly entertaining, like the unexpected theological twist in the Cornell chatbot demonstration. The behaviour of human participants and judges in such tests may also be interesting as a contribution to human psychology, but that's not usually why people discuss the merits of such tests. (However I do recommend teaching students to design chatbots using a sequence of increasingly sophisticated programming constructs, meeting increasingly complex requirements as this can be of great educational value, as explained [below](#).)

Turing himself did not propose a test for intelligence, as a careful reading of his 1950 paper shows, and building machines to do well in such tests does not really advance AI.

So, why did I accept the invitation to be a judge this time? Because I naively thought it might be a good opportunity to sample advances in chatbot developments, on the assumption that only high quality contenders would be selected for the occasion. It turned out that the conditions of the test did not allow much probing, though I gained the impression that the five chatbots I interacted with had not advanced the state of the art much. If you know roughly how most chatbots are programmed, it is not too difficult to come up with questions or comments that the designer has not anticipated and for which a suitable response cannot be concocted by a quick search of the internet. Often a question about physical behaviours of some household material or object in a new context will suffice, e.g. "Will my sock stop these papers blowing about in the wind?" A chatbot with a very large database and flexible pattern matching capabilities may come up with some irrelevant answer referring to windsocks. A human in our culture may be able to produce a clever joke referring to windsocks in reply, but is more likely show common sense understanding of properties of papers, wind, and requirements for paperweights. But no one test is guaranteed to be decisive.

It is clear from Turing's 1950 paper that he did not propose his "imitation game" as a **test** for intelligence, though though he occasionally slipped into calling his non-test a test! What he actually did with the game was set up a prediction about particular capabilities of future computers so that he could demolish arguments that had previously been published against the possibility of machine intelligence. That's why, in previous publications I have referred to [the mythical Turing Test](#).

## Testing a generic, polymorphic design, not an instance.

There is something that is worth testing for the purposes of scientific or philosophical advance, namely a **generic design** that is claimed to produce a variety of partly similar and partly different intelligent individuals, just as the human genome may account for the possibility of many different sorts of humans in different environments, and the elephant genome may account for the possibility of many different elephants with variants of elephant intelligence.

Providing such a generic design, with potentially infinite generative power, because it can be instantiated in a huge class of possible individuals, each of which extends itself in a unique way in response to successive challenges during its lifetime, is comparable to, though much more difficult than, Turing's specification in 1936 of a generic design for computing machinery. Turing's specification for what we now call Turing machines, defining a class of computational systems, was in turn much deeper than any attempt to specify a set of tests for a machine to have computational capabilities. For example, it is impossible to specify a set of behavioural tests that will decide whether something is a universal Turing machine.

(A proof of this is presented in a separate discussion of [black-box tests](#). Martin Escardo informed me that this is actually a special case of Rice's theorem -- which has the consequence that for any 'interesting' computational property it is impossible to determine by behavioural tests whether a Turing machine has that property.)

The question whether some generic design (like the human genome, or elephant genome) has the ability to produce the kind of variety of developmental trajectories that humans (or elephants) are capable of is a much deeper and more interesting task than producing one system that can pass some extended sequence of tests, even over a lifetime.

Moreover it is mathematically impossible to produce a behavioural test that will determine whether any observed individual is an instance of something like the human genome, since over any finite number of tests two very different designs could produce the same behaviours. Although Rice's theorem was not proved in Turing's lifetime (as far as I know) it is clear from the **Mind** 1950 paper that he did not think that 'intelligence' or 'thinking' could be defined precisely and he did not think it sensible to try to devise a test for intelligence or thinking.

By 1936 he had already done something much deeper and more interesting, namely he produced a general theory about properties of a class of machines with computational powers, and a subset with universal computational powers. We still don't have a comparable theory for machines with the powers provided by any of the sophisticated genomes produced by evolution, including genomes for humans, elephants, orangutans, squirrels, or weaver birds. So we lack a theory of the kinds of intelligence that those organisms have.

What I've called [The Meta-Morphogenesis project](#), partly inspired by Turing's ideas, may lead to a collection of such theories, though that will not happen soon.

## **Criticisms and suggestions for improvement of the Turing Test**

Many people have criticised the Turing Test from various standpoints. A common criticism is that is too easy for a machine without human intelligence to pass, on the basis of information fed in advance by a programmer. This often leads to proposals to strengthen the test in some way. One of the best known proposals of this sort is Stevan Harnad's proposal to replace a test using textual interaction in a computer laboratory with a "total Turing test" (TTT) requiring not only conversational abilities but also various types of physical engagement with the environment over extended periods -- possibly a life-time -- demonstrating full human performance capabilities, including, for example, e.g. clearing up and washing up utensils left on a table after a meal (a type of task well beyond current robots and likely to remain so for some time). Harnad presents more detailed discussion than I can summarise here, and at times comes close to my main point, namely that if we wish to answer scientific and philosophical questions, as opposed to achieving an engineering goal and constructing something very like a human being, then it is not enough to simply build one entity that passes some set of behavioural tests, no matter how demanding and how long they take. We need a general theory about the type (or

types) of intelligence we are trying to explain, which covers potentially infinitely many different complete lifetimes and also explains the capabilities of the outliers, like Euclid, Aristotle, Archimedes, Kant, Leonardo da Vinci, Shakespeare, Newton, Bach, Beethoven, Ramanujan, Turing, and others, including rapists, murderers, and humans with various neurodevelopmental abnormalities. No set of behavioural tests is an adequate substitute for a theory about what makes all those capabilities possible (not necessarily simultaneously in any one individual). For more on Harnad's proposals, see Harnad ([1991](#)), and ([2014](#)).

Many more tests are proposed regularly, often ignoring previous proposals. A test proposed recently by Gary Marcus [in his New Yorker blog](#), is an example, namely getting a machine to answer questions or make comments after watching a TV show or online video. Many other "improved" versions of the test have been proposed, usually based on the mistaken assumption that Turing's intention was to propose a test for intelligence.

For reasons given below, and elaborated in a separate discussion of [black-box tests](#) (purely behavioural tests) for intelligence, proliferating tests may be useful for entertainment or engineering purposes, but something totally different from a new intelligence test is required to provide deep answers to scientific or philosophical questions. (As Turing knew: he did not propose a test for intelligence, as his 1950 paper makes clear, though unfortunately most who discuss the test and propose improvements have never read his paper.) Instead of a test for intelligence, we need a [theory](#), and tests for [good](#) theories.

On the basis of the sorts of deep theories covering varied phenomena that Turing himself had already produced (one of which was later presented in his 1952 paper on Morphogenesis mentioned below), I suspect Turing understood how much more important production of a theory, and tests for a theory was than production of tests for a special case of an instance of the theory. However, I suspect Turing had not thought that through when he wrote the 1950 paper.

## **Was the winning chatbot some sort of cheat?**

There are also many reports criticising the 2014 test at the Royal Society because the winning program was designed as a simulation (actually not a very good simulation) of a Russian teenage boy. Critics claim that the judges were told this in advance, making them more likely to be tolerant of errors or odd performance. The claim is that this somehow gave the program an unfair advantage. However, I was a judge and was NOT given the information, and even if I had been I don't think it would have made any difference to my ability to tell that the machine was not a human. Many people who have tried the online version since the London test confirm that it is very easily recognisable as a computer. So all the criticism of the 2014 test based on that feature of the winner are totally irrelevant.

That raises the question how the program was able to fool 10 of the 30 judges. No doubt the organisers of the event will be studying the transcripts, and perhaps questioning the judges about what they did and did not notice. But my main point is that all of this misses the point that there cannot be a good behavioural test for intelligence, just as there cannot be a good behavioural test for computation. I repeat: Turing's paper makes it very clear that he was not proposing such a test.

We need to think about *intelligence* in something like the way Turing had previously thought about *computation*: namely by analysing requirements for various *kinds* of intelligence, including a wide variety of types of animal intelligence, various types of possible machine intelligence, and discussing which kinds of machinery are capable of explaining which sort of intelligence. That requires a deep theory about products of biological evolution, acknowledging that the concept of "intelligence" required has the feature known to computer scientists as "parametric polymorphism" discussed in

more detail [here](#).

## What did I learn about the state of the art?

When I originally accepted the invitation, to take part in the 2014 Turing test event, I did not realise that because of time constraints for the event, each judge would in effect have only two and a half minutes to judge each 'player', since the tests were in five minute slots, using a split screen to allow each judge to interact concurrently with both a human and a machine, in an attempt to distinguish them.

On the day, I felt the time available did not permit me to evaluate progress in chatbot design, though deciding which was the human seemed to me to be very easy in each case. I'll find out later whether I was fooled by any of the chatbots, but I was pretty sure that I managed to identify all five of them by the first, second or third response. However, one of my tests was 'failed' by all the humans as well as all the machines. In response to "My new hearing aids should help us communicate" not one of them pointed out the irrelevance of hearing aids to textual interaction. One gave a one-word response: 'brilliant' which might have expressed pleasure at improved communication or an obscure compliment to the tester. Later responses convinced me that was a human.

Despite the shortage of time, the differences between human and machine responses usually seemed clear. Of course I may be wrong. I'll update this when I have been told how many decisions I got right: However, I don't think that I meet Turing's requirement that the judges should be "average interrogators" (see below). That would rule out someone who had built a (toy) chatbot, namely the Birmingham [Pop11 Eliza](#) (based on a toy chatbot developed about 35 years ago at Sussex university as a teaching demonstration for undergraduates, who played with it then learnt to build simple chatbots of their own, as precursors to deeper work on language understanding). If not being average rules me out as a participant, that increases the proportion of participants who were fooled by any machine I identified!

The short time did not really allow me to probe any of the chatbots in depth, so I was not able to learn much about their strengths and weaknesses. The short time limit was required in order to fit enough separate judging sessions into the time available, though I suspect Turing's reference to five minutes for his "Imitation game" was intended to allow five minutes for each player, especially as most humans (in particular those whom he referred to as "average") are not high-speed typists. His actual words were:

*"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. The original question, 'Can machines think?' I believe to be too meaningless to deserve discussion."*

(His guess regarding memory capacity by the end of the century was remarkably accurate. I suspect that in the year 2000 a higher proportion of the 'average' human population would have been fooled by some of the more sophisticated chatbots than now because since then a lot more humans have learnt about what computers can and cannot do and the sorts of stupidity they often display. But it is too late to test that hypothesis.)

## Educational value of chatbot development

I normally refuse to be involved with Turing Tests, because, for various reasons given in the rest of this document, I regard such tests as having no scientific or philosophical merits, as I assume the organisers knew when they invited me. However, playing with, criticising, and implementing chatbots can be a useful educational activity, enabling students to develop some understanding not only of AI and programming but also of the nature of human language and the differences between pattern-based and syntax-based language comprehension and use, among other things.

That was one of the key teaching ideas when we introduced courses in Artificial Intelligence, including programming, for students in Arts and Social Science subjects at Sussex University, around 1976, led by the late Max Clowes

<http://www.cs.bham.ac.uk/research/projects/cogaff/sloman-clowestribute.html> His ideas about teaching had a deep impact on our teaching of programming and AI:

<http://www.cs.bham.ac.uk/research/projects/cogaff/sloman.beginners.html>

See also the tutorials on chatbots, pattern matching, linking a chatbot to a changeable database -- as in the SHRDLU program in [Winograd \(1972\)](#) -- and related educational topics here:

<http://www.cs.bham.ac.uk/research/projects/poplog/cas-ai/video-tutorials.html>

---

## What's wrong with the Turing "Test"?

Using chatbots for educational and other practical purposes is one thing. Claiming that producing one that can fool a human for five minutes or even for five days or five years is of great scientific or philosophical value is another, especially if the use of vast numbers of pre-stored, appropriately annotated, text fragments reduces the machine's need to understand or replicate how humans generated the utterances that such machines are designed to mimic.

More importantly, we already know how to create things that are able to behave like human beings over many years, in very many different environments, including all the environments in which humans have survived. That's what we do when we produce babies, as many others have pointed out. But we don't know how they work, and we don't know all the design requirements that a typical human (or other intelligent animal) needs to satisfy in order to live a human life.

This illustrates my claim that merely being able to produce a machine that performs over any (bounded) length of time as a human would does not indicate that we have increased our understanding of human intelligence, or any other kind of intelligence. It depends on the intellectual (scientific, philosophical and engineering design) knowledge used in the process. Making babies uses less of such knowledge than making chatbots, but neither requires deep understanding of how human or animal minds work.

For example, we don't really know much about what the functions of human and animal vision are, although many people who think they do know have begun trying to build robots that can satisfy their supposed functional requirements (e.g. segmenting the environment and recognising/labelling the fragments segmented), or taking in binocular 2-D images, or moving images, and creating 3-D models that can be displayed on a screen in 'fly through' mode. My brain cannot do that. I can't even draw static versions of most of the things I can see very clearly. (My wife is much better at this.) So many of the supposed capabilities involved in human or animal vision are inventions of researchers, not based in deep understanding of the functions of biological vision systems.

## Requirements for seeing

Unfortunately the work on vision in AI and robotics that I know about ignores most of the subtle visual functions concerned with detecting, using, and understanding affordances in the environment, discussed in Gibson [1966](#) and [1979](#), and extended in [Sloman 2011](#), and in the online [Meta-Morphogenesis](#) project papers, especially [this discussion](#) of unsolved problems regarding biological and machine vision.

In particular, most, and perhaps nearly all, vision researchers assume that visual systems acquire metrical information and attempt to create information structures in which metrical values are used, or, when there's not enough information for precise metrical inferences, probability distributions over metrical values are constructed instead.

So if your visual system cannot tell whether a gap is exactly 45 cm wide, many researchers assume that instead it produces a sort of internal graph of the probabilities of all the possible values around 45 cm that are consistent with the retinal stimulation. That's a very complex process and if it is being done for all estimates of distance, direction, curvature, gap sizes, speeds, etc. then the requirements for the brain to handle all that probabilistic information sensibly become computationally and mathematically very demanding -- or perhaps totally intractable. This may explain why the competences of current robots seem to be so much more restricted than those of very young children and many other animals.

(I find it particularly mathematically implausible that brains representing spatial structures, relations, and processes of change in structures and relations, in terms of collections of probability distributions could have discovered the beautiful theorems and proofs in Euclidean geometry that were first discovered without the help of mathematics teachers thousands of years ago. No current machine that I know of comes close to that. For examples of the kinds of reasoning required see: <http://www.cs.bham.ac.uk/research/projects/cogaff/misc/triangle-theorem.html> Hidden Depths of Triangle Qualia <http://www.cs.bham.ac.uk/research/projects/cogaff/misc/triangle-sum.html> Old and new proofs concerning the sum of interior angles of a triangle.)

Probability based vision researchers tend to ignore the biologically much more plausible alternative of making use of large amounts of definite (i.e. not merely probable) information about partial orderings (nearer, bigger, more curved, sloping more steeply, heavier, faster, etc.) that provide an adequate basis for a large amount of animal action, especially in connection with servo-control mechanisms (using visual feedback). It is often possible to be absolutely certain that A is taller than B when you can see both, even if you can't estimate the height of either with much precision. Likewise you can often tell with certainty whether your hand is moving away from, or towards an object or neither, even if you cannot estimate the speed of movement or the distance accurately. For more on this see [this web page](#).

The ability to make and use such comparative perceptual judgements and to reason about their consequences is an important aspect of natural vision, and seems to be part of the basis of some human mathematical competences, for example in reasoning that if A is further away than B, and B is further away than C, then A is further away than C, and noticing that this is not an empirical generalisation but can be understood as an example of an inviolable constraint on partial orderings. (It is not at all clear how brains do this or how to give robots the ability to make such mathematical discoveries.) So this is an example of the kind of challenging requirement that needs to be met by a new generic design for a class of spatially competent machines. And explaining how that mathematical reasoning ability might be implemented in either animal brains or future machines is part of the requirement for the kind of deep research that should replace either attempts to improve the (mythical) turing test or make machines that appear to pass the test, e.g. if asked or if given a practical task that requires such

reasoning. More examples of such requirements are concerned with proto-mathematical discoveries made and used by very young children, before they know what they are doing, i.e. "Toddler Theorems": <http://www.cs.bham.ac.uk/research/projects/cogaff/misc/toddler-theorems.html>

Vision researchers in AI/Robotics, neuroscience and psychology mostly ignore the deep connections between human spatial perception and the ability to do mathematics, especially abilities required for making the sorts of geometric mathematical discoveries reported in Euclid's *Elements* over two thousand years ago, which includes many theorems and proofs of the theorems that must have been discovered originally when there were no mathematics teachers. How? Partial answers are suggested in Sloman [MKM08](#) and [AISB-10](#).

Turing was far too intelligent to claim that the sort of ability displayed by a competent performer in his imitation game was adequate for anything like human intelligence, apart from a very narrow subset. His purpose was very different, namely to refute a specific set of arguments other thinkers had produced about the impossibility of machine intelligence.

## **Changing fashions of argumentation against the Turing test.**

To those of us who remember the debates about the Turing test in the 1970s to 1990s, the current spate of discussion about the recent test is an amusing indicator of changes of fashion in modes of thinking and reasoning. For example, the old 'huge lookup table' objection to the Turing Test, that used to be discussed at length in usenet debates and elsewhere seems to have lost its appeal, or perhaps hasn't been noticed by the younger generation of thinkers, whereas a few decades ago, it was often considered an obvious objection to any general purpose behavioural test for intelligence. (A separate discussion of "black-box testing" takes this point further: <http://www.cs.bham.ac.uk/research/projects/cogaff/misc/black-box-tests.html>)

## **The mythical Turing test.**

Part of the explanation for the general lack of understanding of what Turing was getting at is that the majority of those who refer to "The turing test" have not read his 1950 paper.

I have sat through many research seminar presentations that allude to a test for intelligence allegedly proposed by Turing, after which, when asked, the speaker confesses to not having read what Turing actually wrote.

It should be clear to anyone who has read the Mind 1950 paper, that Turing did not propose any test for intelligence. The 1950 paper has been reprinted in many places, most recently in the prize-winning 2013 collection of papers and commentaries, with contents listed [here](#).

The 2013 collection also includes my paper 'The mythical turing test', available also as a 'preprint' that will be revised from time to time [Sloman \(mythical\)](#).

It argues, as this paper does, that Turing was far too intelligent to propose the sort of test that is attributed to him, and that he was merely making a fairly limited prediction about what he thought computers might be able to do by the end of the century. His main purpose was to analyse and refute previously published arguments that seemed to imply that his prediction could not succeed. For now, it's not important whether his arguments worked. The point is that he did not propose a behavioural test for intelligence and that attempting to do so would be misguided because it does not address the deep research problems.

All the proposed variants of the test fail to address the need to identify a design that is based on an explanatory theory rather than a design whose performance merely matches some observed behaviours.

A discussion of limitations of what can be learnt from "black box" tests of turing machines can be found [here](#), including a brief mention of Rice's Theorem (Roughly: no "interesting property" -- in a technical sense of that phrase -- of a computational system  $C$  can be proved by a turing machine observing the behaviour or inspecting the rules of  $C$ ).

## What were Turing's long term aims?

Clues to Turing's long term aims come from a number of interesting unelaborated comments, which I suspect reflect ideas he was working on, including ideas about the importance of chemistry in biological information processing, presented in his highly influential 1952 paper on the "Chemical basis of morphogenesis", also included in the 2013 collection.

One clue about what he might have thought about possible future developments is his aside regarding digital (discrete) computers (in the 1950 paper):

"Strictly speaking there are no such machines. Everything really moves continuously. But there are many kinds of machine which can profitably be thought of as being discrete-state machines."

And this statement made in passing, but not elaborated:

"In the nervous system chemical phenomena are at least as important as electrical."

I suspect those two comments and the examples in the 1952 paper suggest that Turing had started thinking about chemical information processing mechanisms which existed in a variety of organisms long before brains evolved, and continue to play important roles in animal bodies, e.g. fighting infection, repairing damage, and of course growing brains in embryos. One feature of chemical information processes in biological organisms is that they combine continuous change with molecules (or parts of molecules) moving together or apart, folding, twisting, unwinding, etc., with discrete processes such as formation or release of chemical bonds, and many catalytic and autocatalytic processes. It is possible that that combination can do things discrete computers (including Turing machines) cannot do. If so, AI and Robotics in future may have to extend the repertoire of available implementation mechanisms for their designs.

These ideas suggest a long term project of trying to identify major transitions in information processing in organisms, including changes in both what is done and how it is done, e.g. using chemical forms, neural forms, and forms of computation based on use of virtual machinery in more complex evolved species.

I call the project to investigate those evolutionary developments and their consequent developmental (epigenetic) processes the Meta-Morphogenesis (M-M) project, partly because it was inspired by Turing's 1952 paper.

The project includes attempting to specify a variety of forms of biological (human and animal) information processing, e.g. in visual perception, mathematical discovery (especially in geometry and continuous topology), nest building by many animals, including weaver birds, without assuming that they can all be implemented in digital computers. The project is outlined here:

<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/meta-morphogenesis.html>

If important aspects of human intelligence rest on such mechanisms and (a) we still have no deep and broad specification of the requirements to be met, (b) we don't yet know what mechanisms can meet those requirements and (c) we don't fully understand the evolved mechanisms, or the biological functions for which they are essential components, then we may be unable to replicate human-like intelligence in machines, in the foreseeable future.

There are certainly many machines performing impressively with **fragments** of human intelligence, and in some cases **superhuman** fragments, because of the speed and complexity of what they do. But there are also many aspects of human and animal intelligence that we are nowhere near emulating in machines. Examples include the mathematical abilities that must have led to the discoveries eventually collated in Euclid's *Elements* over two millennia ago, and many animal abilities including the weaving of long thin leaves to make hanging nests done by weaver birds, demonstrated in this video:

<https://www.youtube.com/watch?v=6svAIgEnFvw>

---

## More interesting and useful challenges than the (mythical) Turing Test?

[Sloman \(2013\)](#) described a very different kind of test, which I suspect is closer to what Turing might have worked on if he had lived longer, namely a test for a *theory* of intelligence, not a test for an *instance* of what the theory is about.

What Turing had done previously provides clues as to the task he was addressing. In particular, in his ground-breaking work on computation in 1936 he did not provide a 'test for computation' by specifying how each computer should be tested by comparing it with a standard computer, or an 'average' sample of standard computers. (Before then, most computers were human, though there were some mechanical and electrical calculating and sorting devices, and before that there were [Jacquard looms](#). See the [Jacquard Loom Walkthrough](#) in Stacey Harvey Brown's video).

Instead of proposing tests for whether computations are being performed, Turing did something much deeper in 1936. He produced an analysis of a class of competences exhibited by humans when doing arithmetical calculations or logical derivations -- by making successive sequences of marks on a surface, such as pencil marks on paper.

He then proposed an abstract schema, a generic specification, for a type of machine, now known as a Turing machine, that could be instantiated in infinitely many ways (so he was talking about properties of a class of machines, not of any one machine); and he demonstrated that the instantiations covered a very wide variety of sequences of manipulations of numerals and other symbols. By allowing the lengths of the sequences to be arbitrarily long (requiring a potentially infinite tape in the Turing machine) he showed that any such machine had the potential to perform any one of an infinite variety of such computations. In particular any known arithmetical calculation could be translated into a sequence of such operations that a Turing machine could perform.

He also showed, surprisingly, that a subset of the instances, the **Universal** Turing machines, could each model all the (infinitely many) other TMs.

The capabilities emulated included a variety of different forms of symbolic reasoning that mathematicians and logicians had studied, which previously only humans could do.

That work required him to start with a precise specification of the requirements for what humans were able to do, so that he could prove mathematically that all the requirements could be satisfied in Universal Turing machines.

Later 'Universal' proved to be a misnomer because there are wider classes of types of computation (information processing) not covered, and Turing began to explore some of them. For example, when he died he had been working on Chemical mechanisms, reported in the 1952 paper on the chemical basis of morphogenesis and since chemical mechanisms use a mixture of discrete and continuous operations, they may be able to perform important tasks that a purely discrete machine cannot perform. Discrete operations can approximate continuous ones up to a point, but there are notorious unavoidable consequences of 'rounding errors' which in some cases can add up to huge errors. Moreover, Turing machines, and most of the systems about which theoretical computer science deals, are systems whose internal behaviour consists of a succession of discrete states. In contrast biological information processing systems may include continuous processes and generally include many different interacting processes that are not synchronised. Such possibilities have important philosophical implications that have not generally been understood by philosophers or cognitive scientists. See this discussion of "Virtual Machine Functionalism"  
<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/vm-functionalism.html>

## Proposing a test for computation vs proposing a theory

Turing did not propose a collection of behavioural tests for computation, and claim that any machine satisfying those tests was a computing machine. It was very important that he had a way of specifying computing machines that allowed mathematical proofs of what they could and could not do, so that it was not necessary to rely on behavioural tests; though such tests might be needed to ensure that a physical mechanism implementing such a machine was reliable, or worked fast enough to be useful, or had a large enough memory for a particular class of problems.

### Note:

*This comment needs to be expanded with a detailed account of:*

- (a) how a mathematically specified design for a digital mechanism can be proved mathematically to have certain properties;*
- (b) how a design for a physical implementation conforms to that mathematical design specification;*
- (c) how a particular physical machine conforms to its intended physical design. (This may require a great deal of inspection and testing, including separate testing of internal components of the system -- in general behavioural testing of the whole system would be grossly inadequate as a test procedure, as would external detection of evidence of internal electro-magnetic activity. Unfortunately experience of such engineering challenges is not normally part of a degree in philosophy or psychology.)*

Similarly, if we want to understand human intelligence or something more general that includes human intelligence, we need a **generic** specification of a **type of design** that can be instantiated in many ways, with appropriate consequences, something like the human genome being instantiated in many new born babies who grow up in an enormously wide variety of environments and develop many different sorts of intelligence and competence and interests and achievements, etc. This idea was presented in Sloman (2007) and (2010) but in a way that made it hard for readers to understand. (That sort of general design is a special case of what can be produced by the far more general processes of biological evolution by natural selection operating on a sufficiently powerful medium of change, as discussed in the [Meta-Morphogenesis](#) project proposal.)

Simply trying to design one machine and then testing it may be fun, but it's really just "hacking", with little or no scientific or philosophical value, though it may have useful consequences, including educating future philosophers, scientists and engineers about the technology -- and especially what does *not* work: a most important form of learning whose significance is under-appreciated by many teachers. (Good teachers, especially good mathematics teachers understand this.)

I find it very surprising that so many intelligent people take the (mythical) Turing Test project seriously as a way to specify what intelligence is, instead of attempting to specify a class of machines that can develop a wide variety of instantiations of the concept of intelligence. In his 1950 paper Turing, mistakenly in my view, hinted at a way of doing that by building a robot with a large empty memory except for some powerful general learning mechanisms, and then showing how such a robot could learn and develop with help from teachers. Many AI researchers have been seduced by similar ideas, but I think most of them fail to grasp the point made by John McCarthy in [1996](#), namely

Evolution solved a different problem than that of starting a baby with no a priori assumptions. .... Instead of building babies as Cartesian philosophers taking nothing but their sensations for granted, evolution produced babies with innate prejudices that correspond to facts about the world and babies' positions in it. Learning starts from these prejudices. What is the world like, and what are these instinctive prejudices?

I suspect that if Turing had continued the research begun in his 1952 paper on the Chemical basis of Morphogenesis he would have appreciated this point.

## Epigenetic conjectures

What alternative is there to a large empty memory and a powerful learning engine? In an attempt to answer this Jackie Chappell and I tried to generalize Waddington's notion of an epigenetic landscape to permit a landscape that is built as an individual develops partly under the influence of the current environment, and partly under the influence of the environments of ancestors via the genome, using a multi-layered developmental process. For more on this see our [\(2007\)](#) paper.

It should be obvious that the idea of ANY test for intelligence is silly, because there are so many varieties of intelligence, including e.g. weaver birds, though not all are equally intelligent:

<https://www.youtube.com/watch?v=6svAIgEnFvw>

Human infants and 3 year old toddlers can grow up to be professors of quantum physics, composers, bricklayers, hurdlers, doctors, reporters, farmers, plumbers, parents, etc.

Yet all of them would perform poorly on most tests for intelligence in the first few years of life.

However some toddlers who would fail most intelligence tests seem to make mathematical discoveries unwittingly and most adults never notice:

<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/toddler-theorems.html>

The concept of intelligence, like many other philosophically puzzling concepts, exhibits something like the feature known to computer scientists as "parametric polymorphism" (probably discovered much earlier by mathematicians and given other labels). There's a brief tutorial on that here:

<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/family-resemblance-vs-polymorphism.html>

---

**NOTE (CPHC RESPONSES):** The turing test event has produced a huge amount of discussion and criticism on mailing lists by computing academics in the UK, e.g. the CPHC list <http://cphc.ac.uk/>

Here's a sample, not in any significant order, quoted with permission of authors.

- Prof. Ian Gent, at St Andrews University, has some interesting reflections on this episode, and reactions to it, here:  
<http://iangent.blogspot.co.uk/2014/06/why-passing-of-turing-test-is-big-deal.html>
- Some of the academics suggested that it was not appropriate to lend credibility to the type of testing done by commenting publicly, and even that the judges who had been fooled by the winning program had themselves "failed the turing test" (written as a joke, I assume).
- Prof Anil Seth, Sussex university posted this riposte.  
<http://anilkseth.wordpress.com/2014/06/09/the-importance-of-being-eugene-what-not-passing-the-turing-test-really-means/>
- Prof Murray Shanahan, Imperial posted at least three attempts to debunk the Turing Test claim:  
<http://www.buzzfeed.com/kellyoakes/no-a-computer-did-not-just-pass-the-turing-test>  
<http://www.telegraph.co.uk/science/science-news/10888828/Captain-Cyborg-the-man-behind-the-controversial-Turing-Test-claims.html>  
<http://metro.co.uk/2014/06/09/actually-no-a-computer-did-not-pass-the-turing-test-4755769/>
- Prof Austin Tate, Edinburgh University, mentioned a good discussion about the event on Radio 4 on June 11th, and posted a link to a record of his chatbot interacting with itself as another individual: <http://blog.inf.ed.ac.uk/atate/2011/10/09/avatar-to-avatar-chatbots/> He also mentioned the MSc in e-learning at Edinburgh  
<http://blog.inf.ed.ac.uk/atate/2012/11/29/msc-in-e-learning-graduation/> "From 2011 when I did our School of education MSC in e-Learning taught by online methods."
- Prof Jon Crowcroft, Cambridge University, noted that human ideas about the nature of humanity change, e.g. between emphasising rationality and emphasising various forms of irrationality and emotionality, suggesting that standards for passing a test for human like intelligence would need to change.
- Prof Yorick Wilks, Sheffield university and Florida Institute of Human and Machine Cognition, who had submitted a winning program to a previous Turing test competition posted these comments on the recent event:  
<http://www.theguardian.com/commentisfree/2014/jun/11/ai-eugene-goostman-artificial-intelligence>  
"Don't believe the science hype - we haven't created true AI yet"
- Prof Stevan Harnad, Southampton university reminded us of his various papers proposing improved versions of the Turing test, e.g. <http://eprints.soton.ac.uk/252601/> Harnad, Stevan. (2001) Minds, Machines and Turing: The Indistinguishability of Indistinguishables Journal of Logic, Language, and Information (special issue on "Alan Turing and Artificial Intelligence"). Also at: <http://link.springer.com/article/10.1023%2FA%3A1008315308862>
- <http://blogs.lse.ac.uk/impactofsocialsciences/2014/06/10/turing-testing-and-the-game-of-life/> Stevan Harnad (June 10th, 2014) Turing Testing and the Game of Life: Cognitive science is about designing lifelong performance capacity not short-term fooling.

**NOTE** My own work is not primarily an attempt to create "True AI" (whatever that might be) but mainly to understand natural intelligence in its many forms and to answer old philosophical problems about the nature of life and mind. That understanding could not be expressed in a design for a particular kind of chatbot or robot, only in a collection of abstract design specifications and abstract requirement specifications, along with theories about how the designs and the requirements can vary. In principle, such theories could be useful for explaining the competences of many sorts of animals, and accounting for commonalities and differences in information processing across different species. Such a theory could also be used in the production of a wide variety of human-like and other robots with partly shared genomes and partly shared environmental influences during learning and development. In particular, the ideas sketched in [the 2007 paper](#) written with Jackie Chappell might explain some of the commonalities and differences in patterns of development, using a generalisation of Waddington's notion of the "epigenetic landscape" of a species.

## References (To be extended)

- Mary-Ann Russon gave a useful summary of an interview on this topic by phone and email here:  
<http://www.ibtimes.co.uk/why-turing-test-not-adequate-way-calculate-artificial-intelligence-1452120>  
Another journalist completely screwed up what I wrote to him, but later corrected his report and apologised.
  
- <http://www.cs.bham.ac.uk/research/cogaff/05.html#200502>  
Aaron Sloman and Jackie Chappell, The Altricial-Precocial Spectrum for Robots, in *Proceedings IJCAI'05*, Edinburgh 2005, pp. 1187--1192,
  
- <http://www.cs.bham.ac.uk/research/projects/cogaff/07.html#717>  
Jackie Chappell and Aaron Sloman,  
Natural and artificial meta-configured altricial information-processing systems, *International Journal of Unconventional Computing*, 3, 3, 2007, pp. 211--239,
  
- J.J. Gibson, 1966, *The Senses Considered as Perceptual Systems*, Houghton Mifflin, Boston,
  
- J. J. Gibson, 1979, *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston, MA,
  
- <http://www.newyorker.com/online/blogs/elements/2014/06/failing-the-turing-test.html>  
Gary Marcus  
What Comes After the Turing Test?  
New Yorker Online  
June 9, 2014
  
- John McCarthy, 2008 (on his web site since 1996)  
The Well-Designed Child,  
*Artificial Intelligence*, 172, 18, pp.2003--2014,  
<http://www-formal.stanford.edu/jmc/child.html>
  
- <http://www.cs.bham.ac.uk/research/projects/cosy/papers#tr0802>  
Aaron Sloman, 2008,  
Kantian Philosophy of Mathematics and Young Robots, in  
*Intelligent Computer Mathematics*, Eds. Autexier, S., Campbell, J., Rubio, J., Sorge, V., Suzuki, M. and Wiedijk, F.,  
LLNCS no 5144, Springer, pp. 558--573,
  
- <http://www.cs.bham.ac.uk/research/projects/cogaff/10.html#1001> Aaron Sloman, 2010,  
If Learning Maths Requires a Teacher, Where did the First Teachers Come From?,  
*Proc. Int. Symp. on Mathematical Practice and Cognition, AISB 2010 Convention*,  
Eds. Alison Pease, Markus Guhe and Alan Smaill, pp. 30--39, ISBN 1902956931,

- Aaron Sloman, 2011,  
What's vision for, and how does it work?  
From Marr (and earlier) to Gibson and Beyond,  
<http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#talk93>,  
Online tutorial presentation, also at <http://www.slideshare.net/asloman/>
- <http://www.cs.bham.ac.uk/research/projects/cogaff/03.html#200302>  
Aaron Sloman and Ron L. Chrisley, 2003,  
Virtual machines and consciousness,  
*Journal of Consciousness Studies*, 10, 4-5, pp. 113--172,
- <http://www.cs.bham.ac.uk/research/projects/cogaff/07.html#716>  
Aaron Sloman, 2007,  
"Why Some Machines May Need Qualia and How They Can Have Them:  
Including a Demanding New Turing Test for Robot Philosophers,"  
*AI and Consciousness: Theoretical Foundations and Current Approaches*  
AAAI Fall Symposium 2007, Technical Report FS-07-01,  
Eds. A. Chella and R. Manzotti, AAAI Press, Menlo Park, CA, pp. 9--16,
- <http://www.cs.bham.ac.uk/research/projects/cogaff/09.html#910>  
Aaron Sloman, 2010,  
An Alternative to Working on Machine Consciousness,  
*Int. J. Of Machine Consciousness*, 2, 1, pp. 1--18  
(With commentaries and reply by author.)
- <http://www.cs.bham.ac.uk/research/projects/cogaff/11.html#1106c>  
Aaron Sloman, The Mythical Turing Test, 2013, in  
[Alan Turing - His Work and Impact](#),  
Eds. S. B. Cooper and J. van Leeuwen, Elsevier, Amsterdam, pp. 606--611,  
(This is a revised version of a subset of [Sloman 2007](#).)
- <http://www.cs.bham.ac.uk/research/projects/cogaff/misc/meta-morphogenesis.html>  
Constantly evolving overview of the Meta-Morphogenesis project, first described in  
Aaron Sloman,  
'Virtual Machinery and Evolution of Mind (Part 3)  
Meta-Morphogenesis: Evolution of Information-Processing  
Machinery,'  
Eds. S. B. Cooper and J. van Leeuwen,  
*Alan Turing - His Work and Impact*, 2013, pp. 849--856,  
Preprint:  
<http://www.cs.bham.ac.uk/research/projects/cogaff/11.html#1106d>
- A. M. Turing, (1950) Computing machinery and intelligence,  
*Mind*, 59, pp. 433--460, 1950,  
(reprinted in many collections, e.g. E.A. Feigenbaum and J. Feldman (eds)  
*Computers and Thought* McGraw-Hill, New York, 1963, 11--35),  
WARNING: some of the online and published copies of this paper have errors,  
including claiming that computers will have  $10^9$  rather than  $10^9$  bits  
of memory. Anyone who blindly copies that error cannot be trusted as a commentator.

- A. M. Turing, (1952), The Chemical Basis Of Morphogenesis, in *Phil. Trans. R. Soc. London B* 237, 237, pp. 37--72, (and reprinted in the 2013 Elsevier collection.)
- Terry Winograd, 1972, *Understanding Natural Language*, Academic Press
- Bernard P. Zeigler, 1974, A Conceptual Basis For Modelling And Simulation, *Int J General Systems*, 1, 4, pp. 213--228, <http://dx.doi.org/10.1080/03081077408960781>
- 
- 
- 
- 
- 

---

## THANKS

- To the organisers of the Turing Test for giving me the opportunity to take part and have a taste of what current chatbots can do.
- To UK computer science colleagues on the CPHC mailing list who responded to my note on my experiences, and allowed me to quote or refer to their comments.
- To my colleague here in Birmingham, Martin Escardo, who responded to a query by pointing out that my argument for the impossibility of a behavioural test for a Universal Turing Machine was just a special case of Rice's Theorem (summarised and/or discussed on many web sites), e.g. [http://en.wikipedia.org/wiki/Rice's\\_theorem](http://en.wikipedia.org/wiki/Rice's_theorem)
- To Adam Ford for giving me the opportunity to present the arguments in a video interview.
- To Mary-Ann Russon, who patiently accepted an extended tutorial on this subject over the phone and allowed me to see and comment on her draft report before it went to press.
- To Marcin Milkowski for drawing my attention to the important paper by Ziegler (1974) which seems to be very relevant, but which I have not yet studied fully.
- To Harold Thimbleby for useful comments, including comments on the document on [black-box tests](#).
- To Stevan Harnad for drawing my attention to a badly incomplete summary of his views in an earlier version. He has no responsibility for the current summary.

---

[Jump to Table of Contents.](#)

---

Maintained by  
[Aaron Sloman](#)  
[School of Computer Science](#)  
[The University of Birmingham](#)

.

.

.