Aaron Sloman Absolves Turing of
THE MYTHICAL TURING TEST

Aaron Sloman
School of Computer Science, University of Birmingham, UK
`http://www.cs.bham.ac.uk/ axs`

1   Introduction

In his 1950 paper, Turing described his famous 'imitation game', defining a test that he thought machines would pass by the end of the century. For useful surveys of views about the test, see Saygin et al. (2000) and Proudfoot (2011). It is often claimed that Turing was proposing a test for intelligence. I think that assumption is mistaken (a) because Turing was far too intelligent to propose a test with so many flaws, (b) because his words indicate that he thought it would be a silly thing to do, and (c) because there is an alternative, much more defensible, reading of his paper as making a technological prediction, whose main function was to provide a unifying framework for discussing and refuting some common arguments against the possibility of intelligent machines.[1]

I shall try to explain (i) why the common interpretation of Turing's paper is mistaken, (ii) why the idea of a test for intelligence in a machine or animal is misguided, and (iii) why a different sort of test, not for a specific machine or animal, but for a genome or generic class of developing systems, would be of greater scientific and philosophical interest. That sort of test was not proposed by Turing, and is very different from the many proposed revisions of Turing's test, since it would require many instances of the design allowed to develop in a variety of environments. to be tested. That would be an experiment in meta-morphogenesis, the topic of my paper in Part IV of this volume.

2   Turing's 1950 paper

Section 1 of the paper states:

> "I propose to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think.' The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, 'Can machines think?' is to be sought in a statistical survey such as a Gallup poll. But this is absurd. . . . "

---

[1] I have found that many of those who think Turing proposed a test for intelligence, if asked whether they have read the paper, answer 'No'. They simply repeat what others have said. Saygin's and Proudfoot's articles discuss some merits of the test.

Instead of this 'absurd' procedure, he proposes a game, which he calls 'The imitation game', which he uses to formulate a technological prediction:

> *"We may now consider the ground to have been cleared and we are ready to proceed to the debate on our question, 'Can machines think?' and the variant of it quoted at the end of the last section. . . . We cannot altogether abandon the original form of the problem, for opinions will differ as to the appropriateness of the substitution and we must at least listen to what has to be said in this connexion. It will simplify matters for the reader if I explain first my own beliefs in the matter. Consider first the more accurate form of the question.*
>
> *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. . . . "*

The game is not intended to answer the 'too meaningless' question whether machines can think, but enables Turing to formulate his prediction about what can be achieved in 50 years, so that he can discuss several objections to the prediction. Refuting them, by showing that they are all based on unsound arguments, is the main meat of his paper – his way of replacing the 'meaningless' question 'Can machines think?' with a 'relatively unambiguous' question. (We shall see that the question is not as unambiguous as he thought.)

3 Does the test have any value?

Turing's test is far too limited to serve as a criterion for intelligence. Nobody would accept as an employee, or a student, someone whose only known qualification was the ability to fool the 'average' population for five minutes, in 30 per cent of trials. Ability to pass the test is neither sufficient, nor necessary, for being intelligent (or able to think). No engineer would accept 30% success at playing the imitation game for five minutes as either a specification for a worthwhile design or as an acceptance test for a product. Further information would be required, e.g. *how* it managed to do this, under what conditions it succeeded and failed, and whether it used mechanisms that allowed it to overcome its limitations eventually.

Nothing about the test explains *how* a mind can work, or what thinking is. No information scientist would accept Turing's prediction as specifying an explanatory mechanism. By 1950, Turing had already made profound contributions to our understanding of mathematical competence. Passing his shallow test provides no evidence for possession of any such competence. (Interrogation by mathematicians rather than average interrogators might!)

The ability to pass the test could not drive natural selection since it requires the interrogators to have evolved previously. The vast majority of intelligent animals cannot pass Turing's test. Neither can highly intelligent pre-verbal human toddlers. So ability to pass that test is neither necessary nor sufficient for normal animal or human intelligence.

Intelligence is not some unique set of behavioural capabilities: there are different kinds of intelligence (and thinking) evident in nest-building birds, dolphins, elephants, baboons and human toddlers. In the terminology of Ryle (1949), 'intelligence' is a *polymorphous* concept. Its use can vary systematically according to context.[2]

---

[2]Compare what computer scientists call 'parametric polymorphism'.

Though worthless as a test for intelligence or thinking, the imitation game suits Turing's main purpose, namely providing a framework for presenting and refuting a collection of arguments against the possibility of machine intelligence. It has inspired some AI researchers to try to substantiate Turing's prediction, but that has proved difficult. I suspect Turing understood some of the difficulties, unlike some early proponents of AI who rashly predicted the imminent arrival of intelligent machines. Unfortunately, the test has diverted much intellectual effort from a deep study of biological varieties of intelligence and how to model or replicate them.

4   Turing's predictions

Turing was remarkably accurate about the number of bits available in a computer's memory by the turn of the century. His caution in formulating the test (requiring only 30% of testers to be fooled for only 5 minutes) has been justified by the failures of machines to pass the test so far (though they have come close). The failure seems to be a matter of scale rather than the problems of principle that he discussed in his paper. A machine with a sufficiently large and varied collection of stored patterns could obviously pass the test. That's one of the problems with any time-limited behavioural test for intelligence.

Another problem with the test is its dependence on 'average' interrogators. I suspect that by about the year 2000 it actually was possible to fool close to 30% of the 'average' world-wide human population (excluding computer experts and those who had encountered the idea of a Turing test), for up to five minutes. But the 'average interrogator' has changed since then, in ways that Turing did not allow for. Computing technology has continued to advance since 2000, and computers are now doing much cleverer things, while increasing numbers of humans have been learning about what computers can and cannot do, through frequent use, news reports, internet discussions, and so on, making it harder to fool 'average' testers into thinking they are interacting with a human! Many more humans are now able to choose things to say to a machine that may reveal its inability to respond like a human, and the proportion is likely to increase. This relativity to cultural attainment of testers is one of the reasons why the test is so bad as a test of intelligence.

Unfortunately, the misinterpretation of Turing as proposing a test for thinking or for intelligence is so wide-spread that it has led to huge amounts of wasted effort, wasted, because, as Turing himself pointed out, the notion of such a test is based on a question which is 'too meaningless to deserve discussion'.[3] None of this diminishes the value of Turing's main purpose: presenting and demolishing arguments purporting to show that machines cannot successfully play the imitation game.

5   Turing's error about human-like learning

Turing did make one serious error in that paper. In his Section 7, 'Learning Machines', he wrote

In the process of trying to imitate an adult human mind we are bound to think a good deal about the process which has brought it to the state that it is in. We may notice three components.
(a) The initial state of the mind, say at birth,
(b) The education to which it has been subjected,
(c) Other experience, not to be described as education, to which it has been subjected.

---

[3]A related question on which there has been much futile discussion is whether machines can be conscious. I have attempted to write a tutorial introduction to some of the issues and ways of making progress in Sloman (2010c).

Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's? If this were then subjected to an appropriate course of education one would obtain the adult brain. Presumably the child brain is something like a notebook as one buys it from the stationer's. Rather little mechanism, and lots of blank sheets. (Mechanism and writing are from our point of view almost synonymous.) Our hope is that there is so little mechanism in the child brain that something like it can be easily programmed. The amount of work in the education we can assume, as a first approximation, to be much the same as for the human child. ...

Turing, like many AI researchers studying learning machines, grossly underestimated the contribution of biological evolution to the processes of human learning. As John McCarthy McCarthy (2008) put it:
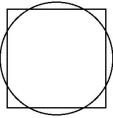
Evolution solved a different problem than that of starting a baby with no a priori assumptions. ... Animal behavior, including human intelligence, evolved to survive and succeed in this complex, partially observable and very slightly controllable world. The main features of this world have existed for several billion years and should not have to be learned anew by each person or animal. In order to understand how a well-designed baby should work, we need to understand what the world is like at the gross level of human interaction with the environment.'

What McCarthy did not point out is that the specification is different for different animals, and different types of machine. Jackie Chappell and I have indicated ways in which such diversity can emerge.[4] This requires much more research in meta-morphogenesis, the topic of my fourth paper in this collection

6 Dichotomies and continua

There is a very common mistake, implicitly made by most who ask: 'Can machines be intelligent?', or 'Can machines think?', namely, assuming that there is a *dichotomy* (a binary distinction) in a complex space where things are very varied. This is as mistaken as assuming there is a binary divide between things that are and things that are not efficient, useful, dangerous, or reliable. Sometimes people who realise that the assumption is mistaken, instead refer to differences of *degree*, e.g. by suggesting that there are differences in *degrees* of intelligence, consciousness, etc. That view makes two related mistakes: (a) assuming that there is a total ordering of cases, as if, for example, species could be put into a linear ordering of animals with more or less intelligence (or consciousness, etc.), and (b) assuming that there is *continuous* variation in kinds of intelligence (etc.).

An example of the first sort of mistake could also be made by a child who finds that some containers can be put inside other containers (e.g. a little box inside a bigger box) and draws the conclusion that all containers form a linear ordering, so that given any two containers Ca and Cb, either Ca can be put inside Cb, or Cb can be put inside Ca, or they are exactly the same size, because their bounding surfaces are the same shape and size.

That's obviously false, because, as shown in the figure, one box may have a square cross section and the other circular, with the diameter of the circle larger than the side of the square and smaller than the diagonal of the square.

---

[4]See Sloman and Chappell (2005), Chappell and Sloman (2007), and Sloman (2008).

There can be many different sorts of competence that are 'dimensions' of intelligence, and different individuals or different species may excel in different dimensions. E.g. one may be very good at designing furniture and terrible at proving theorems in geometry, while the other is good at mathematics and poor at designing furniture. Variations in ability to perceive, learn about, act in, plan in and survive in various types of environment are wide-spread among organisms. Neither species nor individual organisms can be arranged in a linear sequence of increasing intelligence.

The assumption of continuous variation, required for differences of degree, is also false. There are some kinds of knowledge or competence that cannot vary continuously. E.g. learning about arithmetic, or geometry, or grammar, involves learning many distinct items. Competences that are expressed in rules don't have intermediate cases using half a rule, quarter of a rule, etc. Moreover, since genetic makeup is ultimately implemented in chemistry, and since molecules differ discontinuously (e.g. by addition or removal of atoms) it is impossible for species to vary continuously in their genetic makeup (though small differences are possible).

The errors in this section of Turing's paper are related to the mistaken assumption made by some thinkers, that there is some unique competence "general intelligence" which is either present or absent in any individual or species, or is present in different *degrees*, with linear variability along a total ordering.

Similar mistakes can be made about features of machines, such as flexibility, efficiency, reliability: e.g. assuming that there is a binary divide, or a total ordering with continuous variability, for each such feature. The space of machine requirements is not like that, as explained in (Sloman, 2007).

It follows from the above that the very idea of a Turing test or any other test for intelligence is muddled if there is no binary divide between things that are and things that are not intelligent, only a vast variety of cases. Attempting to replace a binary classification with a *measure* of intelligence is similarly mistaken in assuming that there is a total ordering of types of intelligence and, possibly also in assuming that there is continuous variation so that degrees of intelligence can be represented by real numbers.

If there is no total ordering, only a complex space of combinations of competences (just as there is a complex space of combinations of atoms, making the notion of ordering molecules along a line of increasing 'chemicality' (?) misguided), then, instead of using a number or label for degree or amount of intelligence, we need to find ways of *describing* types of intelligence in terms of the combinations of competences that they include, just as we describe chemical molecules in terms of the different combinations of atoms and chemical bonds and also the different properties that result from those structures (e.g. acidity, alkalinity and many more used in drug design).[5].

If varieties of intelligence vary in something like the ways in which sentence structures do, with different components combined hierarchically with varying relationships between the components, then it may be more useful to search for a *grammar* for types of intelligence than a measure. A grammar for types of intelligence might be a specification of varieties of combinations of competences of many sorts that could be implemented in a unified working architecture, and could include combinations that are able to grow themselves, in ways that are modified by their interactions with the environment (as described in Chappell and Sloman (2007)).

---

[5]Added after publication: Similar ideas are presented in (Donald, 2002)

7 What is a grammar?[6]

The notion of "grammar" used here is not restricted to production rules for generation of strings. E.g. during the 1960s, many researchers attempted to generalise Chomsky's notion of a "generative grammar" to include any specification of a class of structures composed from some well defined set of primitives arranged in accordance with permitted relationships, to form new structures that could also be arranged in accordance with permitted relationships, and so on. If the rules are recursive then an infinite variety of structures could be characterised in this way. "Web grammars", for example, were developed to characterise 2-D graphical structures found in pictures. See also (Kaneff, 1970).

From this viewpoint understanding a complex structure conforming to such a grammar amounted to constructing a specification of the structure's derivation from the rules of the grammar, just as parsing a sentence involves constructing a specification of the derivation of the sentence from the grammatical rules. However, in the case of picture grammars the derivation need not have the form of a tree: it could be a graph with cycles, just as pictures can include cyclic structures (e.g. polygons).

A different sort of "grammar" might explicitly characterise ways of generating structures from some initial structures. For example, a grammar for legal chess-board configurations would be a specification of an initial state of the board, plus a specification of permitted changes to any possible state. In this case a derivation of a chess position would be a description of a legal game leading to that position. A quite different notion of a grammar for a chess configuration would characterise pieces and groups of pieces in terms of their positions and relationships. So notions like "pinning" and "forking" and a move being forced would be part of the characterisation of the board structure.

A grammar for chemical molecules might specify ways in which a finite set of atoms can combined in accordance with the laws of physics to form a single structure, and could include intermediate structures that are neither atoms nor molecules, e.g. ions.

In many cases ambiguity is possible: the same structure may have more than one "parse" if there are two or more ways in which the 'rules' allow it to be constructed.

8 What sort of test would be worth while?

If we are to propose tests of the general sort that people take the (mythical) Turing Test to be, namely a test for something being intelligent, or human-like, we'll need to distinguish testing *a particular individual* from testing *a theory about a type of design* for working systems.

It is clear that there are many very different human beings and also that they all share a large collection of common features. We really should be testing a theory about what's common, where the differences come from, and what the implications are. Compare: if a theory about the weather is able to explain only how a particular tornado works and no other weather phenomena, then it cannot be a good explanation of the particular case. Likewise a theory of what's going on when oil burns that says nothing wood burning, or coal burning, or gas burning cannot be a good theory about oil burning.

Or suppose someone claims to have a theory about how to solve algebraic equations and an online computer test that demonstrates the theory. The implemented algorithm solves only quadratic equations: give it any quadratic equation and it will produce the solutions – including complex solutions where appropriate, as required for solving: $X \times X = -1$

---

[6]This section, added at the suggestion of Damien Duff, was removed from the published version because it made the paper too long. I have also used the word "grammar", in the generalised sense explained here, in connection with the idea of a grammar for emotions (Sloman, 1982), as a counter to shallower forms of description of emotions.

Would that be taken as a good test for a theory of equation solving? We would rightly demand something more *generic*.

What would the required sort of *generic* theory of intelligence look like? The closest answer I can give is something like a parametrised specification for a highly polymorphic design for a working system, which can be given different parameters to produce instances of the design, where the instances will be very different in the way that, for example, humans in different cultures, or who talk different languages, or who grow up to have very different competences and interests are different, and yet be as similar as different humans are, a requirement that is very complex and very demanding, and not yet specifiable in detail since we don't yet know enough about what typical humans are like (e.g. how their vision systems work, how they learn, what mechanisms are involved in their motivational and other affective states and processes.)

The parameters, instead of all being supplied at the time the instance is created, would have to be picked up at various times during the development and testing of the instance.

In particular, in order to really understand human intelligence, we should be able to specify a *type* of system, with many different instances, differing as much as humans in different physical and social environments do. For example, as a result of educational and environmental influences, and some individual personality features, instances of the machine would 'grow up' to be philosophers with very different views, including views on what machines can or cannot do, e.g. some becoming like Alan Turing, others like John Searle, or Tom Nagel, or David Chalmers, or Dan Dennett, and perhaps even some like me, since I disagree with all the others![7]

References

Chappell, J., & Sloman, A. (2007). Natural and artificial meta-configured altricial information-processing systems. *International Journal of Unconventional Computing*, *3*(3), 211–239. (http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0609)

Donald, M. (2002). *A mind so rare: The evolution of human consciousness*. W W Norton & Company Incorporated. Available from
`http://books.google.co.uk/books?id=Zx-MG6kpf-cC`

Kaneff, S. (Ed.). (1970). *Picture language machines*. New York: Academic Press.

McCarthy, J. (2008). The well-designed child. *Artificial Intelligence*, *172*(18), 2003-2014. Available from `http://www-formal.stanford.edu/jmc/child.html`

Proudfoot, D. (2011). Anthropomorphism and AI: Turing's much misunderstood imitation game. *Artificial Intelligence*, *175*(5–6), 950–957. Available from
`doi:10.1016/j.artint.2011.02.002`

Ryle, G. (1949). *The concept of mind*. London: Hutchinson.

Saygin, A., Cicekli, I., & Akman, V. (2000). Turing Test: 50 Years Later. *Minds and Machines*, *10*(4), 463–518. Available from
`http://crl.ucsd.edu/ saygin/papers/MMTT.pdf`

Sloman, A. (1982). Towards a grammar of emotions. *New Universities Quarterly*, *36*(3), 230–238. Available from
`http://www.cs.bham.ac.uk/research/cogaff/81-95.html#emot-gram`

Sloman, A. (2007). *A First Draft Analysis of some Meta-Requirements for Cognitive Systems in Robots* (No. COSY-DP-0701). Available from

---

[7]For more on this see Sloman (2010a,c).

    `http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0701`
    (Contribution to euCognition wiki, with help from David Vernon)

Sloman, A. (2008). *Evolution of minds and languages. What evolved first and develops first in children: Languages for communicating, or languages for thinking (Generalised Languages: GLs)?* (Research Note No. COSY-PR-0702). Birmingham, UK. Available from
    `http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0702`

Sloman, A. (2010a). An Alternative to Working on Machine Consciousness. *Int. J. Of Machine Consciousness*, *2*(1), 1–18. Available from
    `http://www.cs.bham.ac.uk/research/projects/cogaff/09.html#910`

Sloman, A. (2010b, August). How Virtual Machinery Can Bridge the "Explanatory Gap", In Natural and Artificial Systems. In S. Doncieux & et al. (Eds.), *Proceedings SAB 2010, LNAI 6226* (pp. 13–24). Heidelberg: Springer. Available from
    `http://www.cs.bham.ac.uk/research/projects/cogaff/10.html#sab`

Sloman, A. (2010c). Phenomenal and Access Consciousness and the "Hard" Problem: A View from the Designer Stance. *Int. J. Of Machine Consciousness*, *2*(1), 117-169. Available from
    `http://www.cs.bham.ac.uk/research/projects/cogaff/09.html#906`

Sloman, A., & Chappell, J. (2005). The Altricial-Precocial Spectrum for Robots. In *Proceedings IJCAI'05* (pp. 1187–1192). Edinburgh: IJCAI.
    (http://www.cs.bham.ac.uk/research/cogaff/05.html#200502)

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*, 433–460. ((reprinted in E.A. Feigenbaum and J. Feldman (eds) *Computers and Thought* McGraw-Hill, New York, 1963, 11–35))