Proceedings of the Towards a Comprehensive Intelligence Test (TCIT): Reconsidering the Turing Test for the 21st Century Symposium,
Aladdin Ayesh, Mark Bishop, Luciano Floridi, and Kevin Warwick (Ed.), at the AISB 2010 convention, 29 March – 1 April 2010, De Montfort University, Leicester, UK

Don't Improve the Turing Test, Abandon It

Drew McDermott[1]

Abstract. This workshop is a bad idea, at least if held as part of an AISB meeting. When Turing suggested replacing the question, "Can a machine think?" with the question, "Can a machine pass such-and-such a test?," he never intended for the replacement to be a permanent institution, or for it to form part of a proposed definition or even inductive test for intelligence. Some philosophers may have based an entire subfield of philosophy of mind on this misunderstanding, but AI as a field need not follow them, or respond to them.

1 Intro

"The aim of this symposium is to reconsider the Turing Test ... with the goal of outlining a new test ... that may more usefully be employed to evaluate 'machine intelligence' ..." [1]. I will argue that any symposium with this aim is a bad idea, because the Turing Test is a bad idea. Although this point has been argued eloquently already by Hayes and Ford [6], I will make the case from a slightly different angle. I will argue that Turing's intentions were misunderstood, and that it is only the insecurity of researchers in cognitive science that has kept endless discussions about his proposal on the table.

2 Turing Would Have Repudiated Turing-Style Tests

Alan Turing never doubted that people were machines [9]. This view may have stemmed from the oddness of his personality; the thought that he was a machine may have come more naturally to him than to the average person [12]. From the beginning, when he thought about computational issues, even before computers actually existed as physical entities, he assumed that similar things were going on inside human brains and the computers he imagined. Initially the analogy ran from brains to machines: In his seminal paper [21] on what are now called Turing machines, in thinking about what machines could calculate, he worked from idealized models of what a person following an algorithm could calculate, breaking the process down into the simplest steps we can conceive of [21, sect. 9]. But by the 1940s, the process was going the other way. Now the physical Turing machines did exist. He had access to the Manchester machine [8, 23], and even though its computing capacity seems laughable today, he assumed, correctly, that technology would change the boundary conditions dramatically and soon. Now he wanted to know when computers would acquire abilities comparable to their creators'.

Like everyone else, he made the mistake of assuming that sensory and motor skills would be relatively straightforward to decipher, so that the difficult part of AI would be intellectual skills such as mathematics and playing chess. Chess was one area in which it was possible even in the late forties to envision competing against humans, so Turing worked on a chess program, although it was never fully implemented.

However, he was impatient with frontal attacks on the AI problem, and wanted to get to the heart of the matter quickly. Hence the speculative paper "Computing Machinery and Intelligence," which appeared in *Mind* in 1950.[2] In the first paragraphs of this paper, Turing suggests a strategy for speculation:

> 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

[1] Computer Science Department, Yale University, New Haven, CT, USA email: drew.mcdermott@yale.edu

[2] Turing's predictions for how easy different tasks would be, and how powerful computers would be, come from this paper as well.

> 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. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words. (p. 433)

What follows is a description of the Imitation Game, in which an interrogator (labeled C) attempts to decide which gender to assign to two participants (labeled A and B) whom C can communicate with only via teletype (or "chat" in today's parlance). The man is supposed to attempt to convince the interrogator that he is the woman. Turing concludes by saying

> We now ask the question, "What will happen when a machine takes the part of A in this game?" Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, "Can machines think?"

It is somewhat confusing whether he intended for the machine literally to take the place of a man playing against a woman with the judge deciding which was the woman, or whether he assumed an implicit *mutatis mutandis* clause, so the "man–woman" issue was replaced by "machine–person." I will assume the latter, which is the standard reading.

Notice the way Turing invites us to consider his game. Realizing that it is impossible to define "machine" and "think" in a way that will not beg the question "Can a machine think?," he proposes to "replace" it by a "closely related" question. He never explains what this "close relation" is.

Later in the paper, reviewing progress, he summarizes thus:

> It was suggested tentatively that the question, "Can machines think?" should be replaced by "Are there imaginable digital computers which would do well in the imitation game?" (p. 442)

But "replaced" for how long? To what end? Later he shows his hand considerably more clearly:

> 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. Nevertheless I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted. (p. 442)

The question requiring definition of "machine" and "think" is, according to Turing, "meaningless." I believe his reason may be found in his earlier remark that answering it would require a "Gallup poll" on what the words meant. By the end of the 20th century, he predicts, when computers can pass the Turing test (a prediction that misfired), "opinion" will have changed so that the outcome of the poll would be different, and one could casually speak of machines thinking as a normal thing. Turing and AI would have achieved an end-run around the question by focusing on winning the Imitation Game as a more tangible research goal.

Does all this show that Turing meant to *define* intelligence (or "thinking") as the ability to pass the Turing Test? It certainly doesn't seem so. The "close relation" one question bears to the other at the beginning of the paper has become rather more distant by the end.

The closest Turing comes to equating "thinking" with passing the Turing Test is in his reply to the "argument from consciousness," which he quotes from Geoffrey Jefferson's Lister Lecture of 1949: "Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain…" [7, p. 1110]. Note that the conclusion Jefferson rejects is that "machine equals brain," a conclusion Turing makes clear (p. 451) he does not accept either. The title of the section of Jefferson's paper in which these arguments appear is "Thinking," a word Turing didn't want to use. So it's odd that Turing quotes Jefferson for the argument from consciousness. Perhaps his choice is explained by the fact that, just like Turing, Jefferson is tempted to equate thoughts and words — a temptation for *any* intellectual. Just before the quote Turing extracted, Jefferson says that the "schism" between computing machines

and brains "arises over the use of words and lies above all in the machines' lack of opinions, of creative thinking in verbal concepts" [7, p. 1110]. It sounds to me like the two mens' positions were not that far apart, Jefferson might indeed have abandoned his position if confronted with a machine that could carry on a dialogue such as the one Turing pictures refuting him [22, p. 446].

In any case, Turing's response is clumsy. He actually gives *two* arguments against Jefferson, in rapid succession:

1. "According to the most extreme form of [Jefferson's] view the only way by which one could be sure that machine thinks is to *be* the machine and to feel oneself thinking. One could then describe these feelings to the world, but of course [according to Jefferson] no one would be justified in taking any notice."
2. "Likewise according to this view the only way to know that a *man* thinks is to be that particular man. It is in fact the solipsist point of view. It may be the most logical view to hold but it makes communication of ideas difficult. A is liable to believe "A thinks but B does not" whilst B believes "B thinks but A does not." instead of arguing continually over this point it is usual to have the polite convention that everyone thinks."
(Turing [22, p. 446], emphasis in original)

The second argument is a rather casual restatement of a standard argument against solipsism: Since it would be impossibly awkward for everyone to walk around acting like a solipsist, let's agree that all people "think," meaning that they are conscious. Turing tries to slip machines in by invoking "the polite convention that *everyone* thinks," but to allow "everyone" to include machines begs the question — the question of consciousness, that is.

The first argument is more novel. It is that, if Jefferson is wrong, it would be intensely frustrating to be a machine and have people, for apriori reasons, refuse to believe that one could be frustrated. Out of simple charity one should give a machine the benefit of the doubt if it could express what it sincerely *believed* to be feelings. Elsewhere in the paper Turing refers to the Muslim "belief" that women have no souls,[3] and says that it seemed to him, as a nonbeliever, that believers must accept that God could give any being a soul if he wanted to. It is not too far-fetched, in the image of the suffering machine, to sense Turing's own frustration at being unable to talk about his homosexuality in terms his society could, psychologically and theologically, accept.

As everyone knows all too well, Turing did not survive his own silent suffering. If he had, we would have had ample time to hear from him on what he really intended by proposing the Turing Test. I think he would be shocked to find a symposium in whose description we read, "[The Turing Test] has become synonymous as 'the benchmark' for Artificial Intelligence in popular culture," but "[d]uring a recent debate at AISB09 . . . , it became clear that, whilst the Turing Test has served well in driving the research agenda in AI forward, it should no longer serve as a true test of intelligence" [1]. There are two reasons to disagree sharply with these statements.

First, the Test has not served particularly well in driving our research agenda forward, and if he were alive today Turing would almost certainly *not* be organizing a team to win the Loebner Prize [19]. Instead, the AI community has worked on a series of much narrower problems, which is disappointing to those hoping for "general artificial intelligence" to crystallize suddenly. In the nineteenth century, visionaries may have fantasized about someday understanding biology well enough to create life, but the only one who actually tried was the fictional Frankenstein. For the real scientists, it was a triumph to synthesize urea, even if this achievement was only a step along a long road, whose twists and turns could not be imagined until it was actually followed.

Second, no AI researcher should accept that the Turing Test has served as "a true test for intelligence." (I picture robots being brought before an examiner who would administer the True Test; those who failed it would perhaps be consigned to the assembly lines. It sounds so medieval, like seeing if witches can float.) Intelligence in a field is the ability to out-think one's competitors in that field; intelligence in a tennis player is the ability to fool their opponent into guessing wrong about where they are sending the ball. A "true test" for this kind of intelligence is a tennis match. I meet the objection that this leaves us unable to test for "general intelligence" by replying that I doubt there is such a thing. If we are interested in conversational intelligence about some topic, then a test might be whether someone who requires information or help about that topic prefers to talk to competitor A rather than competitor B. If we are interested in conversation about nothing in particular, then the question is whether a person would rather talk to competitor A or competitor B. Or a judge might look at a dialogue between A and B and decide which one had made wit-

[3] It is a myth that Muslims believe this, although their actual beliefs about women's souls aren't very palatable either [20].

tier use of their "opponent's" utterances. (Some future program might compete against a future Dorothy Parker or Oscar Wilde.)

In any case, a test of intelligence is always a test of "more intelligent than" rather than simply "just as intelligent as." As Hayes and Ford point out [6, p. 973], putting it the latter way makes success depend on inability to detect a difference: "This is such a fundamental error that it has been given a title: confirming the null hypothesis."

I think Turing would have made clear that he never intended his imitation game as an institution, a *permanent* replacement for substantive questions about thinking. He just wanted to talk about playing this game as a more productive enterprise than talking about "thinking" as things stood *in 1950*. It is tragic that he was foreclosed from clarifying his intentions in later years because he was hounded to death by prosecution, hormone injections, and public humiliation. This tragedy is a burden the AI community has had to bear, one of many little and not-so-little injuries homophobia has inflicted on humanity over the centuries. Unfortunately, it looks like we're going to be bearing it until we either succeed in creating a machine that will realize Turing's vision and make quibbles about whether it thinks seem silly, or admit that it can't be done.

3 Philosophy and Cognitive Science

It is a puzzle why an essentially *philosophical* question about improving a test we don't need should be discussed at a *scientific* conference. It raises the more general question about what the proper role of philosophy should be within cognitive science, or, more generally, what the relation is between philosophy and science. It is a commonplace that sciences split off from philosophy as they *become* sciences. Physics ceased being philosophy in the seventeenth century; psychology, in the twentieth. Chemistry split off from both alchemy and philosophy at the same time, in the late eighteenth.

In Restoration England, there was serious debate about whether getting your hands dirty with experiments was the right way to understand the physical world, versus apriori reasoning within frameworks such as classical geometry [18]. The experimenters won the debate, and ever since philosophers can't sit at the science table unless they have an experiment to propose or a piece of mathematics to contribute.

That's a bit harsh. You could also argue that philosophers have a job to do in clearing up conceptual confusions. For example, philosophers such as Hilary Putnam rightly cried foul at the widespread complacency that the Copenhagen Interpretation brought to the interpretation of quantum mechanics [13] for decades. It might be thought that philosophers' many contributions to the literature on the Turing Test are just part of an effort to help us understand it better.

And yet, when you look at all the papers on the Test (see [11] for a recent sampler), the overwhelming majority them seem either to present scenarios in which some entity passes it but is not intelligent after all, or to contain counterarguments regarding those scenarios. The skeptical papers are inspired by the conviction that whatever underlies human intelligence, it can't be computation. Hence the desire to find a way to prove apriori that computation can't pull the weight it's supposed to. Alas, such apriori arguments work only if you have a definition of intelligence. If intelligence is a natural phenomenon, then it can't be defined, any more than "malaria" can be. Before anyone knew what the natural history of malaria was, the word "malaria" just meant, "this disease with characteristic symptoms whose cause we suspect to be bad air" (hence the name). We can now identify malaria with the disease caused by a certain parasite transmitted through mosquitos. But this isn't a *definition*, any more than water is *defined* to mean "H_2O."

Turing was sophisticated enough to avoid trying to define intelligence, or even give a sufficient condition for it. But others have fallen into this trap, including AI textbook authors such as Ginsberg [5]: "AI is the enterprise of constructing a physical symbol system that can reliably pass the Turing Test" (p. 17). As soon as enough people had misinterpreted Turing, the literature just grew from there.

This has happened more than once. One example is the odd history of the "frame problem." This was originally a technical issue of representing change using predicate calculus [10]. However, it was misunderstood, initially by Dennett [3], as having something to do with collecting all and only relevant information to reason with, and soon, under Fodor's magisterial supervision [4], it had mushroomed into the problem of scientific induction! Other authors define it differently, in a myriad of ways [2, 14], because the philosophers' version is not a real problem, just a good screen on which to project a set of vaguely defined objections to the possibility of a mere algorithm being able to do as good a job as people do at collecting all and only the relevant information to consider in a given circumstance. Anyone who is interested can now look up the answer to the actual frame problem [15, 16], but the philosophers' frame problem has taken on a life of its own. (For an excellent overview of both, see [17].)

Philosophers can, of course, do what they please. The question is why cognitive scientists should pay them special attention. Is there some research program in AI that is fundamentally misconceived for reasons philosophers can explain? Specifically, is there some problem connected to the Turing Test that bears on the current research agenda of AI? Of course not.

4 Conclusions

There is no need to find an improved version of Turing's test for intelligence, because this test as it is conceived of today is largely the creation of philosophers for their own purposes. Turing certainly never intended his thought experiment to provide a synonym for "intelligence" (or "thinking"), or even to provide a practical way of testing for the presence of intelligence. He was writing for an audience in 1950, when the idea of machines that could think seemed preposterous. In today's conditions his thought experiment appears to be redundant — just as he predicted it would be. It's time to lay it to rest.

REFERENCES

[1] Aladdin Ayesh. Call for papers, AISB 2010 symposium on topic: Towards a comprehensive intelligence test (TCIT): Reconsidering the turing test for the 21st century, 2009.

[2] *The Frame Problem in Artificial Intelligence*, ed., Frank R. Brown, Morgan Kaufmann, San Francisco, 1987.

[3] Daniel Dennett, 'Cognitive wheels: the frame problem in artificial intelligence', In Pylyshyn [14].

[4] Jerry Fodor, 'Modules, frames, fridgeons, sleeping dogs, and the music of the spheres', In Pylyshyn [14].

[5] Matthew L. Ginsberg, *Essentials of Artificial Intelligence*, Morgan Kaufmann Publishers Inc, 1993.

[6] Patrick Hayes and Kenneth Ford, 'Turing Test considered harmful', in *Proc. Ijcai*, volume 14, pp. 972–977, (1995). Vol. 1.

[7] Geoffrey Jefferson, 'The mind of mechanical man', *Brit. Med. J*, **1**(4616), 1105–1110, (1949).

[8] Simon Lavington, *A History of Manchester Computers, second edition*, The British Computer Society, Swindon, 1998.

[9] David Leavitt, *The Man Who Knew Too Much: Alan Turing and the Invention of the Computer*, W.W. Norton & Co., Atlas Books, New York, 2006. "Great Discoveries" series.

[10] John McCarthy and Patrick Hayes, 'Some philosophical problems from the standpoint of artificial intelligence', in *Machine Intelligence 4*, eds., Bernard Meltzer and Donald Michie, 463–502, Edinburgh University Press, (1969).

[11] James H. Moor, *The Turing Test: The Elusive Standard of Artificial Intelligence*, Kluwer Academic Publishers, Dordrecht, 2003. Studies in Cognitive Systems.

[12] Henry O'Connell and Michael Fitzgerald, 'Did Alan Turing have Asperger's syndrome?', *Irish J. of Psychological Medicine*, **20**(1), 28–31, (2003).

[13] Hilary Putnam, 'A philosopher looks at quantum mechanics', in *Mathematics, Matter and Method. Philosophical Papers* (2nd edition), ed., Hilary Putnam, Cambridge University Press, (1979).

[14] *The Robot's Dilemma: The Frame Problem and Other Problems of Holism in Artificial Intelligence*, ed., Zenon Pylyshyn, Ablex Publishing Co, Norwood, N.J., 1987.

[15] Raymond Reiter, *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems*, The MIT Press, Cambridge, Mass., 2001.

[16] Murray Shanahan, *Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia*, MIT Press, 1997.

[17] Murray Shanahan, 'The frame problem', in *Stanford Encyclopedia of Philosophy*, (2009). the.

[18] Steven Shapin and Simon Schaffer, *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*, Princeton University Press, 1985.

[19] Stuart Shieber, 'Lessons from a restricted Turing Test', *Comm. ACM*, **37**(6), 70–78, (1994).

[20] Jane I. Smith and Yvonne Y. Haddad, 'Women in the Afterlife: The Islamic View as Seen from Qur'an and Tradition', *Journal of the American Academy of Religion*, **43**, 39–50, (1975).

[21] Alan Turing, 'On computable numbers, with an application to the *entscheidungsproblem*', in *Proc. London Math. Society* , volume 2-42, pp. 230–265, (1937).

[22] Alan Turing, 'Computing machinery and intelligence', *Mind*, **49**, 433–460, (1950).

[23] Alan Turing and R.A. Brooker. *Programmers' Handbook (2nd Edition) for the Manchester Electronic Computer Mark II.*, 1952.