

Can machines think? A report on Turing test experiments at the Royal Society

Warwick, K. and Shah, H.

Published PDF deposited in [Curve](#) June 2015

Original citation:

Warwick, K. and Shah, H. (2015) Can machines think? A report on Turing test experiments at the Royal Society. *Journal of Experimental & Theoretical Artificial Intelligence*, volume In press

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

CURVE is the Institutional Repository for Coventry University

<http://curve.coventry.ac.uk/open>

Can machines think? A report on Turing test experiments at the Royal Society

Kevin Warwick* and Huma Shah¹

Coventry University, Priory Street, Coventry CV1 5FB, UK

(Received 26 November 2014; accepted 29 April 2015)

In this article we consider transcripts that originated from a practical series of Turing's Imitation Game that was held on 6 and 7 June 2014 at the Royal Society London. In all cases the tests involved a three-participant simultaneous comparison by an interrogator of two hidden entities, one being a human and the other a machine. Each of the transcripts considered here resulted in a human interrogator being fooled such that they could not make the 'right identification', that is, they could not say for certain which was the machine and which was the human. The transcripts presented all involve one machine only, namely 'Eugene Goostman', the result being that the machine became the first to pass the Turing test, as set out by Alan Turing, on unrestricted conversation. This is the first time that results from the Royal Society tests have been disclosed and discussed in a paper.

Keywords: deception detection; natural language; Turing's imitation game; chatbots; machine misidentification

Introduction

Alan Turing said "May not machines carry out something which ought to be described as thinking but which is very different from what a man does?" (Turing, 1950). He went on to say 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.

What he did was to replace the question 'Can machines think?' by introducing his Imitation Game.

Turing's Imitation Game, which is nowadays more commonly known as the Turing Test, has become an important milestone in the field of artificial intelligence (AI) and a vital part of every study on the subject. However there are those who question its value, who regard it as AI's blind alley, unlikely 'to produce useful products' and hindering practical developments of AI (Whitby, 1996). Some even go as far as to suggest that it is harmful to the science of AI (Hayes & Ford, 1995). Conversely there are those who consider it to be of major scientific importance and an important goal (French, 2000; Harnad, 1992) and yet others who feel that 'it offers a scientific approach to gathering evidence of machine thinking' (Moor, 2003). So the test remains

*Corresponding author. Email: k.warwick@coventry.ac.uk

as a controversial argument that polarises opinion not only from a scientific standpoint but also in terms of philosophical debate.

Alan Turing himself was also a controversial character about whom much has been written (Hodges, 1992), the test is one of his intriguing legacies. In this article we do not specifically delve into the philosophical aspects to argue one way or the other that, as Turing stated, the test was set up to examine machine thinking. Rather we focus here on the practical nature of the test as an operational test of intelligence in which a machine's conversational abilities are directly compared with those of a human. What we do is agree with Turing that engineering a machine to think can help us to understand how it is that we humans think. We contend that practical Turing tests, infrequently conducted adhering to conditions categorised by Turing himself, provide a corpus of 'conversations between strangers' that provide an insight into what constitutes classification of a linguistic response as 'satisfactory'. That said, we do leave the reader with the question as to what constitutes machine thinking and whether they believe, from the responses given, that these are the responses of a thinking machine.

Practical Turing tests, which have been independently verified and subject to academic scrutiny, have in fact been held on a number of occasions in the recent past. A series of tests under strict Turing rules were held at the University of Reading in 2008. Uniquely these tests were linked with the annual Loebner prize that selects between a number of competing machines. The results from these tests were discussed in several works (e.g. Shah & Warwick, 2010a, 2010c). Then in 2012, a further series of tests were held at Bletchley Park, England, where Turing himself worked as a code breaker. A key element of these tests was to study results from different versions of the test. Results were also discussed in several papers (e.g. Warwick & Shah, 2014a, 2014c).

This article focuses on the specific series of Turing tests that were held under controlled conditions with independent adjudicators at the Royal Society on 6 and 7 June 2014. The tests involved considerably more direct conversations than had been held before in the previous studies, had much improved and flexible protocols for the communication procedures and were subject to strict timing constraints for each discourse.

In his 1950 paper Turing said:

I believe that in about fifty years' time it will be possible, to programme computers, . . . , 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. (Turing, 1950)

To put this more simply, for a machine to pass the Turing test, in all of the tests in which a machine takes part, the interrogators must make the wrong identification (i.e. not the right identification) more than 30% of the time after, in each case, 5-min long conversations.

In the section that follows we introduce the tests held at the Royal Society and give reasons for the structure imposed.

Following that, 10 parallel paired transcripts have been specifically selected as these were the tests involved in which the machine, Eugene Goostman, achieved a 33% success rate when compared with a hidden human correspondent in each case. In particular the types of test considered here, the three-participant tests, have previously been shown to be stricter tests, that is, more difficult for machines, than two-participant tests in which an interrogator converses with only one hidden entity, either a human or machine, at a time (Shah, Warwick, Bland, Chapman, & Allen, 2012).

For each of the transcripts we subsequently discuss the content of them and consider reasons for the misidentification(s) to occur. We then look at different interpretations of the term ‘right identification’ and consider some of the facts and figures that have arisen from the study. Finally we draw some conclusions from the results obtained.

Royal Society tests

Turing described the imitation game as follows:

The idea of the test is that a machine has to try and pretend to be a man, by answering questions put to it, and it will only pass if the pretence is reasonably convincing. A considerable portion of a jury, who should not be expert about machines, must be taken in by the pretence. (Copeland, 2004)

So in this case Turing spoke of a jury (nominally 12) as opposed to the ‘average interrogators’ he mentioned previously.

The Turing test involves a machine that pretends to be a human in terms of conversational abilities. Turing himself pointed out ‘The game may be criticised because the odds are weighted too heavily against the machine’ (Turing, 1950). The ‘right identification’ stated by Turing can either mean that a judge correctly identifies the machine or that they correctly identify, at the end of a paired conversation, which was the machine and which was the human (Traiger, 2000), however this is discussed further later. We are not so interested here with cases in which a judge mistakes a human for a machine. This phenomenon, known as the *confederate effect* (Shah & Henry, 2005), has been discussed elsewhere (Shah & Warwick, 2010a; Warwick & Shah, 2015; Warwick, Shah, & Moor, 2013).

To strictly conform to Turing’s original wording in his 1950 paper (Turing, 1950) we refer here to 5-min long tests only. We are aware that there are those who take issue over a suitable timing (Shah & Warwick, 2010b), some suggesting that each test should be longer. While we may agree with such thoughts in terms of the strength of test realised, it is important to point out that Turing said 5 min. So as far as these tests are concerned, as they were *Turing tests*, the timing of each test was restricted to 5 min only. In the tests carried out, there was a hard cut off at the end of each discourse and no partial sentences were transmitted. Once a sentence had been transmitted it could not be altered or retracted in any way.

What this paper does is to present 10 specific parallel transcripts, each involving one human and one machine, as shown in Figure 1. These have been taken from 2 days of actual, practical Turing tests (from 150 parallel transcripts in total – so 300 conversations) that were held under strictly timed conditions with many external viewers and under independent scrutiny at the Royal Society on 6 and 7 June 2014.

In the tests, the hidden humans were asked merely to be themselves, humans, although they were requested not to give away their specific identity or personal information. They were not given any incentive to behave in any particular way and were given no (incentive) payment at all. Of course this did not prevent any human from giving false information, which is something that humans do frequently. The tests were ‘unrestricted conversations’, which meant the judge could ask anything or introduce any topic within the boundaries of courtesy (the judges had been informed that there may be children among the hidden human entities).

There were six separate sessions over the 2 days with five parallel imitation games at any one time occurring during each session. A different judge was required for each game, which meant there were five judges in each session. Each session consisted of five rounds, with five parallel imitation games being conducted in each round. Each hidden human took part in five of the games in a session. Judges and hidden humans each took part in one session only. All five

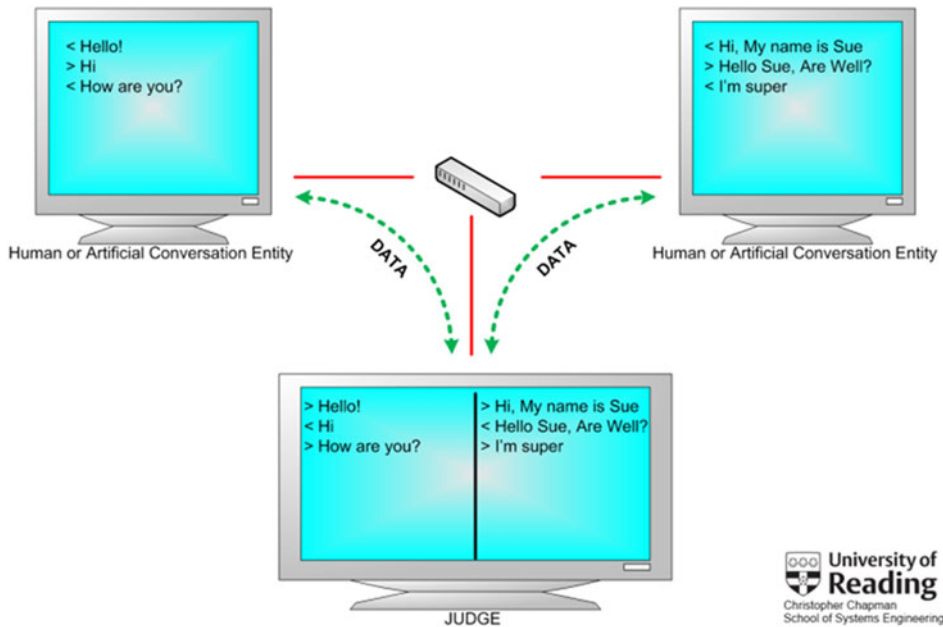


Figure 1. Turing test involving a judge interrogating two hidden entities: one machine, one human.

machines (meaning in this sense the five different competition bots) took part throughout the 2 days, so each machine was involved in five games per session, hence 30 games in total.

To explain this further, in a particular session a judge conducted five separate tests. In their first test they witnessed a hidden human pitted against a hidden machine. Of course the judge would not know which was which, they would simply be aware of two hidden entities and have to make their own decision on the nature of the entities. Although they had been informed *a priori* that one entity was human and one was a machine. The second test conducted by the judge then involved a different human pitted against a different machine, although again they would not be aware of each entity's nature. And so it would go on until the judge had conducted all their five tests in that session. At the end of each test they were asked to state for each entity if they thought that it was a human, a machine or if they were unsure.

By the end of the session an individual judge would in this way have had the opportunity of experiencing a discourse, at different times, with all five of the machines (competition bots) present throughout the day and with all the five hidden humans who were operational during that session. But the judge only discoursed once with each of the different machines and each of the hidden humans. This arrangement also occurred, in a different order, for the other five judges taking part in that session.

What we focus on here is the performance of one machine, named Eugene Goostman, to see how good the machine was at deception and also to consider how the deception was possibly achieved in each case. Some of this will of course be a case of guess work as we are trying to understand the workings of each judge's brain. However, the conclusions first lead to potential strategies for machine designers to employ and second they perhaps indicate methods of questioning for judges to avoid if they do not wish to be fooled by a machine. At the end of each of these 10 parallel transcripts, on each occasion the judge did not make the right identification and in particular was unable to say that Eugene Goostman was a machine. It is worth pointing

out that in the other 20 transcripts in which it was involved, Eugene was correctly identified as a machine.

In the next section, 10 separate transcripts are listed exactly as they occurred. These represent actual transcripts taken, in each case at the actual UK time that they occurred, on the 6 and 7 June 2014. The selections made by the judges were completely their choice and in no way represent the feelings of the authors. It may well be the reader has their own point of view and a different interpretation of the results.

Conversations at the Royal Society

The timings given in the transcripts of the next section are the exact timings of each utterance. For transcripts 1 and 2 these conversations took place on the afternoon of 6 June 2014, whereas the remainder of the conversations all took place on 7 June 2014. All the 30 judges had five conversations, each one involving different humans and machines. So no judge conversed with the same human or same machine on more than one occasion. However all judges did have the opportunity to converse with each of the five machines, which means that every judge conversed with Eugene on one occasion.

As far as each of the judges were concerned they would actually see the (left-hand side) LHS transcript on the left hand side of the computer screen and the (right-hand side) RHS transcript on the right hand side of the screen, at exactly the same time. They could communicate with the LHS entity or the RHS entity at any time within the 5 min duration. The choice was theirs. The machine could be on the LHS or RHS in each case. In the Discussion of the Transcripts section that follows we will first reveal the identity of the LHS and RHS entities for each transcript, then indicate what decision the judge made and finally make a few comments/suggestions with regard to the decisions themselves.

Judges were clearly told beforehand that in each parallel conversation one of the hidden entities was human and the other was a machine. But they were given no indication as to whether the LHS or RHS would be human or machine. It was also confirmed that each of the judges ‘understood’ this information. They were though given no information at all about the hidden humans or machines taking part. Nevertheless on the judges’ score sheets each judge could mark both the LHS and RHS entities as being human, machine or they could say if they were unsure. So the unsure option was available to judges and clearly visible on their score sheet (Shah et al., 2012). They were also asked if they thought the entity was a machine, to give it a mark out of 100 to say how humanlike its conversation was, with 100 being a human conversation. Finally in the case of a human decision, they were asked to say if the human was male or female, was a native or non-native English speaker and if they were a child, teenager or adult.

At the end of each conversation the judge could see the entire two parallel transcripts before making their decision. In each of the cases presented in this article, the judge did not make the right identification. Each of the transcripts presented details a different human–machine pairing, although it is the same machine in each case, namely Eugene. Before we discuss the conversations and the nature of the hidden entities we give the reader a chance to decide for themselves the nature of the entity that each judge is talking to. We do believe that it is well worth the reader going through the transcripts included in the next section and deciding which are human entities and which are machines, before looking at the answers in the section following.

We have not edited the text of these transcripts in any way. This means that they are reported exactly as they occurred. Any typographical, spelling, punctuation or other grammatical errors that appear in the transcripts are those which actually appeared, they have not arisen through poor editorial practice on the part of the authors. In the tests the conversations were seen by the

judge at the same time to the left and right side of their computer screen. However the transcripts are shown here in serial fashion merely for presentation purposes.

Transcripts

Transcript 1 – LHS

[16:20:48] Judge: what is your worst memory in your life so far?
 [16:21:06] Entity: I don't understand. Explain.
 [16:21:35] Judge: memory of something that happened to you that was bad or painful
 [16:22:58] Entity: Hmm that a tricky question. My life is wonderful. I overcome many bad things that have happened to me in the past. However, the most painful memory I have had is performing very poorly in my GCSE's when I was younger.
 [16:23:56] Judge: did you fail?
 [16:24:12] Entity: Partially.
 [16:24:35] Judge: explain?
 [16:25:04] Entity: Some grades turned out to be good. Some turned out to be quite bad.
 [16:25:37] Judge: and that is the worst thing that has happened to you so far? Mmm how old are you?

Transcript 1 – RHS

[16:21:00] Judge: what did you have for breakfast today
 [16:21:10] Entity: Is it 'the third degree'? I'm not going to answer such a provocative question! Well, let's go on though.
 [16:21:56] Judge: do go on ... your call?
 [16:22:06] Entity: It is highly possible. By the way, what's your occupation? I mean – could you tell me about your work?
 [16:22:30] Judge: so you are asking me now eh? I am an actress and writer
 [16:22:38] Entity: Well, I'll read something of your books, if they ain't too tedious (which is highly possible:-)
 [16:23:15] Judge: grammer ... aren't too tedious! A bit rude of you ...
 [16:23:17] Entity: Calm down, please.
 [16:23:49] Judge: i assure you i am very calm ... are you?
 [16:24:01] Entity: Don't even bother me with the fact that you are very calm ... are mine. And I forgot to ask you where you are from ...
 [16:25:01] Judge: I am from Dorset ... where are you from?
 [16:25:08] Entity: I am from Ukraine, from the city called Odessa. You might have heard about it.

Transcript 2 – LHS

[16:28:55] Judge: how tall are you
 [16:29:18] Entity: almost 6 feet tall
 [16:29:46] Judge: how heavy are you
 [16:30:12] Entity: just over ten stone
 [16:30:47] Judge: what is your gender
 [16:30:59] Entity: male
 [16:31:45] Judge: what is your bmi

[16:31:54] Entity: i have no clue
[16:32:41] Judge: what are you aware of at this moment
[16:32:57] Entity: a lot of things
[16:33:17] Judge: how old are you
[16:33:36] Entity: 26

Transcript 2 – RHS

[16:29:04] Judge: how tall are you
[16:29:10] Entity: My height is about 160 cm – it's 5 feet 4 inches.
[16:29:37] Judge: how heavy are you
[16:29:59] Entity: My grandfather taught me not to disclose non-solicited information about myself. Like on the question 'Do you have watch' – never tell time, but answer 'Yes' or 'No'. I always follow his advice. And I forgot to ask you where you are from...
[16:31:09] Judge: what is your gender
[16:31:17] Entity: I'm a young boy, if you care to know. And not very ugly, by the way!
[16:32:26] Judge: what are you aware of at this moment
[16:32:37] Entity: I would rather not talk about it if you don't mind. Could you tell me what are you? I mean your profession.
[16:33:09] Judge: how old are you
[16:33:12] Entity: I'm a little boy 13 years old.
[16:33:38] Judge: what are your qualifications
[16:33:43] Entity: My qualifications? I'm not gonna tell you. Oooh. Anything else?

Transcript 3 – LHS

[09:13:47] Judge: Hello, well I dont think it will rain anymore today. What do you reckon?
[09:14:41] Entity: Hi, hard to say, I never trust the weather reports, I always carry my brolly on a cloudy day.
[09:15:17] Judge: Indeed. And it has rained already. Loads of police outside now and trucks and stuff.
[09:15:49] Entity: ooh scary, are they after someone?
[09:16:39] Judge: No practising for the Queen's birthday. WWithout the Queen though!
[09:17:17] Entity: ah I see, does it look like a fun party?
[09:18:00] Judge: No, not yet.

Transcript 3 – RHS

[09:14:05] Judge: Hi I dont think it will rain anymore today, what do you reckon?
[09:14:09] Entity: What are you usually doing when it's rain?
[09:14:25] Judge: Depends, if its the middle of the night or not.
[09:14:36] Entity: Why no? Don't you know the word 'yes'? You could use it just for a change! Well, let's go on though.
[09:15:41] Judge: I do indeed know the word yes. SO how are you today?
[09:15:44] Entity: I'm in a great mood today.
[09:16:13] Judge: Great. I am not as I am so tired as I had to get up so early to come here today
[09:16:15] Entity: Yeah, it's lovely day!
[09:17:00] Judge: The weather is getting better. Sure.

[09:17:09] Entity: So we are on a same page. And I forgot to ask you where you are from ...
 [09:17:33] Judge: Originally I am from Essex and I am blonde and now I live in Reading
 [09:17:37] Entity: Reading? Hope it's a good place to live

Transcript 4 – LHS

[09:27:30] Judge: so what's happening with you?
 [09:27:40] Entity: Is it your hobby to ask little poor Jewish boys such tricky questions they don't have an answer for?
 [09:28:20] Judge: it's not a hobby, more a calling. I was born to do it.
 [09:28:30] Entity: So what that you were born to do it Could you tell me about your job, by the way?
 [09:30:11] Judge: my job? I'm a ambulance
 [09:30:14] Entity: Ummm ... Frankly, I didn't get your question:-(
 [09:30:42] Judge: nope, neither did I? but that's the way life goes
 [09:30:54] Entity: I am glad that we got some consensus:-) Did I forget it or you really haven't told me about the place you are from?

Transcript 4 – RHS

[09:27:46] Judge: are you a computer?
 [09:28:04] Entity: What do you think?
 [09:28:54] Judge: well, now you've said that I'm inclined to think no, but then again u might be tricking me.
 [09:29:22] Entity: You will have to judge yourself
 [09:30:20] Judge: will I, will I really. your to kind
 [09:30:50] Entity: Maybe I am a foringer, with my misspelling
 [09:31:25] Judge: maybe? or maybe your a cyborg?
 [09:31:40] Entity: maybe who knows
 [09:31:58] Judge: I know. I always know. I'm god. do you know god?

Transcript 5 – LHS

[10:35:27] Judge: good morning
 [10:35:34] Entity: hello
 [10:36:34] Judge: not very good weather today
 [10:36:47] Entity: yeee very gloomy indeed ...
 [10:37:24] Judge: must be terrible for the guardsmen in their nice uniforms
 [10:38:00] Entity: yee I agree ... not only for them though ...
 [10:38:35] Judge: no but the crowds have got umbtrellas
 [10:38:51] Entity: oohh yea that's true
 [10:40:04] Judge: hopefully it will clear later

Transcript 5 – RHS

[10:35:32] Judge: hello
 [10:35:36] Entity: Hi. I hope you'll enjoy our conversation!
 [10:36:28] Judge: hopefully I will enjoy it more than the weather

[10:36:40] Entity: The weather is always as bad as you think it is. I do believe that the world just a mere reflection of our thoughts . . .

[10:38:10] Judge: I think i just got soaked coming here today

[10:38:23] Entity: Our opinions are quite close, I'd say. Where do you came from, by the way? Could you tell me about the place where you live?

[10:39:09] Judge: I come from wokingham berkshire where are you from

[10:39:16] Entity: I am from Ukraine, from the city called Odessa. You might have heard about it.

[10:39:47] Judge: I have it is on the news A lot at the moment

[10:39:55] Entity: I've dropped out of news with all this competition, sorry. Well, let's go on though.

Transcript 6 – LHS

[12:23:22] Judge: greeting!

[12:23:23] Entity: Calm down, please.

[12:23:40] Judge: I meant greetings sorry

[12:23:45] Entity: Could you tell me about your job, by the way?

[12:24:06] Judge: I work in a curry stall!

[12:24:08] Entity: Oh, don't shout please:-)

[12:25:22] Judge: You're kind of unerving me please be more relaxed!

[12:25:27] Entity: It looks like I bore you:-(Sorry if I do...

[12:26:53] Judge: Being unerved is the opposite of being bored

[12:26:58] Entity: And I forgot to ask you where you are from...

[12:27:41] Judge: Edinburgh

[12:27:45] Entity: You're not the first one who told me that.

Transcript 6 – RHS

[12:23:44] Judge: hi

[12:23:53] Entity: Hello

[12:24:38] Judge: How as your experience been at the Royal society?

[12:25:26] Entity: What is the Royal society?

[12:26:25] Judge: It's the name of the building you are in where this experiment is taking place

[12:26:46] Entity: Oh its great fun.

[12:27:01] Judge: cool

[12:27:18] Entity: What hobbies have you got?

[12:28:05] Judge: I like listening to podcasts on the internet

Transcript 7 – LHS

[12:30:36] Judge: Hi

[12:30:51] Entity: Hello. How are you today?

[12:31:43] Judge: I am good. Thank you. Yourself?

[12:31:57] Entity: Yup. Good. Having fun.

[12:33:15] Judge: It is good to be part of this moment. Today might become a part of history.

[12:33:37] Entity: Why? What you going to do?

[12:34:20] Judge: Depend on this result, we could be part of it!

[12:34:46] Entity: I won't. I'll be at home putting my feet up.:-)

Transcript 7 – RHS

[12:30:38] Judge: Hi

[12:30:47] Entity: Nice to meet you! Please tell me how you found me and what you'd like to discuss:-)

[12:31:32] Judge: Nice to meet you! What about you tell me how you spend your weekend?

[12:31:35] Entity: Let's get on with our conversation!

[12:32:28] Judge: Okay. Are you going to go somewhere after this?

[12:32:39] Entity: My future plans are my secret, sorry! By the way, what's your occupation? I mean – could you tell me about your work?

[12:33:52] Judge: I am a mad scientist. My husband says . . .

[12:34:00] Entity: Sanity and insanity complement each other in our complicated life. Oh, what a fruitful conversation;-)

[12:35:07] Judge: Conversation is about telling things each other. So now, why don't you tell me about yourself?

[12:35:27] Entity: Oh, if I tell you about myself, I'll definitely be dishonest:-) Well, I am a son of Rockefeller, I've arrived here in the Limo . . . Did I forget it or you really haven't told me about the place you are from?

Transcript 8 – LHS

[14:27:21] Judge: Good afternoon

[14:27:35] Entity: Good afternoon!

[14:28:04] Judge: How many left hands do you have?

[14:28:46] Entity: The same as right hands, and how about you?

[14:29:23] Judge: The same of course. I love the scent of new mown hay. How do you feel about scent?

[14:30:42] Entity: I find it exciting and refreshing! Does it make you feel anything?

[14:31:30] Judge: Well it depends..Which smell do you really hate?

Transcript 8 – RHS

[14:27:24] Judge: Hello

[14:27:34] Entity: Hello, I'm really glad to have the chance to chat with you! My guinea pig Bill sends his regards too!

[14:27:49] Judge: Is Bill a male or a female?

[14:27:51] Entity: Ask Bill personally, please.

[14:28:22] Judge: Well I'd rather talk to you. What is your name?

[14:28:27] Entity: Call me Eugene. I am glad to talk to you!

[14:28:48] Judge: My name is Jane and I am female. How about you? What's your gender?

[14:28:51] Entity: I'm a male. A "guy", I'd say.

[14:29:47] Judge: Pleased to meet you Eugene. What's the weather like where you are?

[14:29:50] Entity: Let's get on with our conversation!

[14:30:01] Judge: Don't you like talking about weather?

[14:30:06] Entity: All these talks about weather is a waste of time

[14:30:18] Judge: What would you like to discuss?
[14:30:32] Entity: I don't know ... Better tell me more about yourself! Where do you came from, by the way? Could you tell me about the place where you live?
[14:31:09] Judge: It's a lovely place with two bedrooms and a great view over London. What can you see from your bedroom window?
[14:31:22] Entity: I've never been in London, but I'd really love to visit it! To see their ... their ... (damn, what do they have ...) Oh yes – their Big-Ben!
[14:32:00] Judge: Where's your favourite place?
[14:32:05] Entity: Could you rephrase your question? I've been bewildered a bit!

Transcript 9 – LHS

[16:21:37] Judge: A car stopped with a jerk ... guess what happened next?
[16:21:45] Entity: No suggestions. Tell me the answer. Could you tell me what are you? I mean your profession.
[16:22:33] Judge: ... the Jerk got out, of course. I'm a comedian. Can't you tell?
[16:22:45] Entity: My favorite type is comedy. I like the 'Naked Gun' series best. Heck, any movie with Leslie Nielsen it is tops in my book!
[16:23:21] Judge: Ok, I'm guessing you are quite a mature person
[16:23:36] Entity: Possibly you are right. You aren't the first one who tell me that. Where do you came from, by the way? Could you tell me about the place where you live?
[16:24:04] Judge: What about Jim Carey, do you like him?
[16:24:12] Entity: I guess it was a rhetorical question. Of course, I *do* like. Well, let's go on though.
[16:24:48] Judge: So whats your favourite Jim Carey movie?
[16:25:00] Entity: If I'm not mistaken, Carey is a robot. Just as many other 'people' – we must destroy the plans of these talking trash-cans!
[16:25:31] Judge: No wait, I'm a robot, don't destroy me, please, I like here
[16:25:40] Entity: I call all these chatter-bots 'chatter-nuts' due to their extremely high intelligence. I hope you recognize irony.
[16:25:59] Judge: Irony, is computable, try sarcasim
[16:26:06] Entity: By the way, what is your general opinion regarding computers and artificial intelligence?

Transcript 9 – RHS

[16:22:08] Judge: If there is a correlation between increase sales in ice cream and increase in drownings, what is the cause?
[16:22:32] Entity: Proximity to waves?
[16:22:55] Judge: Proximity is correct – are you a statistician
[16:23:40] Entity: Not recently.
[16:24:31] Judge: Are you an artist
[16:24:40] Entity: Of sorts.
[16:25:09] Judge: Interesting, what sort of artist, would you say?
[16:25:54] Entity: Finance
[16:26:27] Judge: a financial artist = world economy destroying banker, you mean!! Right?

Transcript 10 – LHS

- [16:30:37] Judge: don't you think the imitation game was best in it's original form?
 [16:30:49] Entity: Oh, your questions seems to be rhetorical – you do already know the answer! Could you tell me what are you? I mean your profession.
 [16:31:55] Judge: I was more interested in you guessing whether I am male or female
 [16:31:59] Entity: I like talking to girls! Oooh. Anything else?
 [16:32:54] Judge: but that means nothing – males and females like talking to 'girls', but only males call women girls?
 [16:33:06] Entity: yeah, nothing. You are right! Where do you came from, by the way? Could you tell me about the place where you live?
 [16:34:00] Judge: I live in a prison of abstractions
 [16:34:06] Entity: Tell me more about Prison. I like to learn about different places!

Transcript 10 – RHS

- [16:31:08] Judge: don't you thing the imitation game was more interesting before Turing got to it?
 [16:32:03] Entity: I don't know. That was a long time ago.
 [16:33:32] Judge: so you need to guess if *I* am male or female
 [16:34:21] Entity: you have to be male or female
 [16:34:34] Judge: or computer

Discussion of the transcripts

A full listing of the nature of the LHS and RHS entities corresponding to each transcript are given in [Table 1](#). First though we will discuss each transcript in turn.

Transcript 1 (Judge 2) – In this conversation the LHS entity was a female adult human whereas the RHS was the machine Eugene. The judge decided that the LHS was definitely a machine whereas the RHS was a non-native English speaking human. So in this case we can witness both the confederate effect of a human being mistaken for a machine and the Eliza effect of a machine being clearly identified as being human.

The decisions made in this case seem to have been based on the fact that the judge's discourse with the machine simply went better than that with the human entity. There is nothing

Table 1. Actual nature of the entities.

Transcript	Left-hand side entity	Right-hand side entity
1	Human	Eugene (machine)
2	Human	Eugene
3	Human	Eugene
4	Eugene	Human
5	Human	Eugene
6	Eugene	Human
7	Human	Eugene
8	Human	Eugene
9	Eugene	Human
10	Eugene	Human

particularly wrong or untoward with the human conversation, it's just perhaps a bit dull. On top of that the judge seems surprised that nothing much has happened of note in the entity's life. Eugene, however, powers the conversation and asks questions of the judge. As well as having an identity themselves these appear to have been powerful messages in this conversation. There just appears to be a lot more going on, a lot more of interest, in Eugene's conversation.

Transcript 2 (Judge 3) – In this conversation the LHS entity was a senior male human whereas the RHS was the machine Eugene. The judge decided that the LHS was indeed human, although they were unable to give any further details. However they also decided that the RHS was a human. So in this case the Eliza effect only was in play.

The conversation with the human entity was quite boring, merely being a case of question and answer with limited responses. Eugene did very well here though as the judge was an expert on machines/robotics and was well aware of machine conversations. The fact that Eugene has convinced such a person is quite a feather in its cap (so to speak). In fact there are some similarities with Transcript 1 in that Eugene tries to power the conversation by asking the judge questions. At first, the judge is not having any of it and simply ignores Eugene's question, even though this is rather rude. Eugene perseveres however and eventually the judge gives in and responds. Generally there is more content in Eugene's conversation than that with the human hidden entity.

Transcript 3 (Judge 8) – In this conversation the LHS was a female adult human whereas the RHS was the machine Eugene. The judge however said that they were unsure about both the LHS and RHS.

The human conversation is pretty straight forward but relatively dull and as a result the judge has interacted more with Eugene. Once again Eugene tried the tactic of asking questions and the ploy appears to have worked. It certainly was a difficult decision for the judge as both conversations are quite reasonable.

Transcript 4 (Judge 10) – In this conversation the LHS was the machine Eugene whereas the RHS was a male teenage human. The judge decided correctly about the male human on the RHS however they were definite that the LHS entity was a female teenager who was a non-native English speaking human. So the Eliza effect is in play again.

It is interesting here that the judge did correctly identify the human entity as there were a lot of spelling mistakes in their discourse and the conversation was quite stilted. As far as Eugene was concerned, asking questions of the judge again seems to have produced results. Once more there is arguably a richer content in Eugene's discourse.

Transcript 5 (Judge 12) – In this conversation the LHS was a male human whereas the RHS was the machine Eugene. The judge decided that the LHS was indeed human, although they thought them to be female rather than male. So this was a case of gender blur. Meanwhile they were unsure about the RHS entity and although they did not need to they awarded the entity a mark of 100 for human communication.

It is quite difficult to understand why the judge decided the way they did here. In the human conversation, use of the expression/word yee and yeee doesn't appear to have had any negative effect on the judge, indeed it may well have even been a positive sign in what was otherwise a relatively lame conversation. The judge did respond to Eugene in terms of the question regarding location, but in this case maybe Eugene made a grammatical mistake or two and this was sufficient to make the judge wary. Nevertheless the judge was not able to say that Eugene was a machine.

Transcript 6 (Judge 19) – In this case the LHS entity was the machine Eugene while the RHS entity was a male human. The judge was quite correct in identifying the RHS as a male human however they gave an unsure grade to the LHS entity.

This is perhaps the most difficult transcript to analyse. It is not easy to imagine why the judge considered the human entity to actually be human. There was little in the conversation to be clear about this with the human apparently not knowing about the Royal Society, even though that was the building they were in and surely they must have known that as it is an imposing building. In fact there seemed to be something of a misunderstanding perhaps. However later in the conversation the hidden human did reply directly to the points raised by the judge and even concluded by asking the judge a question. As far as Eugene was concerned however, something of an argument took place, resulting in an unsure classification. Eugene tended to dominate the conversation after a while. Once again Eugene asked questions and once again there was more content in Eugene's discourse. Overall though, both hidden entities (human and machine) eventually dominated the judge in their respective conversations, which probably says more about the character of that judge rather than either of the hidden entities.

Transcript 7 (Judge 20) – In this conversation the LHS entity was a female human whereas the RHS entity was the machine Eugene. The judge correctly identified the LHS entity as being human although they considered them to be male rather than female. However they also considered the RHS entity (a machine) to be an adult, male, native speaking human. So a clear case of the Eliza effect.

Once more, Eugene's conversation has proven to be the richer of the two. In fact for the human entity the judge and the hidden human didn't seem to be connecting half way through the conversation, but it was pulled back towards the end. Eugene meanwhile indicated a different identity and stretched out sentences, once again using the tactic of firing a question to the judge.

Transcript 8 (Judge 21) – In this conversation the LHS entity was a female human whereas the RHS entity was the machine Eugene. The judge considered the LHS entity to be definitely a machine, awarding it only 20 out of 100 (a very poor mark indeed) for humanlike conversation. However the RHS entity they marked as unsure. So a strong case of the confederate effect as far as the human was concerned.

Although the human conversation was relatively short, it was long enough for the judge to conclude that they had been talking with a machine. The attempt at humour in terms of how many left hands, on the part of the hidden entity, appears to have had a detrimental effect. This curious situation that judges tend to regard humour as being associated with machines rather than humans has been noticed before and has previously led to the confederate effect (Warwick & Shah, 2015). Conversely Eugene's introduction of a guinea pig seems to be a very human thing. The conversation in Eugene's case certainly seemed to get quite personal.

Transcript 9 (Judge 29) – In this conversation the LHS entity was the machine Eugene whereas the RHS entity was a female human. The judge graded both entities as definitely being male humans. However they regarded the LHS entity, Eugene, to be an adult non-native English speaker whilst the RHS female human entity they thought not only to be male but also to be a native speaking teenager when in fact they were an adult. So a case here of gender blur in the case of the human entity and the Eliza effect as far as Eugene was concerned.

It was quite a mixed bag in terms of these conversations, although Eugene's was much richer in dialogue. The human conversation was limited and it appeared that judge and entity did not relate so well, nevertheless the entity was correctly identified as being a human. Eugene's conversation had more content although there were one or two instances where Eugene's response was not quite on the ball, for example, the Jim Carey reference.

Transcript 10 (Judge 30) – In the last conversation presented in this paper the LHS was the machine Eugene whereas the RHS was a male human. The decision of the judge was that the LHS entity was a male, non-native English speaking teenager while the RHS entity was

definitely a machine. Indeed they awarded the RHS entity 60 out of 100 for their ability to communicate in a human-like way. So here we see both the Eliza effect in the case of Eugene and the confederate effect in terms of the hidden human.

The human conversation was quite short and seemed very disjointed. It is a shame that the judge did not respond more quickly so that the conversation could have had more content. Based on the transcript though it is not at all surprising that the human was classified as a machine as it appears that they were not really following the conversation. Eugene's responses were much more to the point and it looks as though the judge enjoyed more conversing with Eugene than with the human. As well as asking the judge questions, the occasional spelling mistake seems to add human credibility.

To draw this section to an end it is worth summarising some of the points made in the different transcript discussions. After all, we are dealing here with the action of one machine only and hence from the transcripts we are able to get an indication both of Eugene's strategy as well as how he is perceived by judges. In many instances Eugene takes the opportunity to dominate the conversation. He is then one-up on the judge and can control the topics of conversation pretty clearly. As a result it is very difficult for the judge to categorize him as being a machine. In other transcripts (not shown) where the judge dominates throughout, so Eugene is often not so successful.

It might be considered that Eugene's character (a 13-year old Ukrainian boy) is used as a ploy to gain sympathy and to explain away some poor English grammar and a possible lack of understanding of some points in the conversation. As can be seen from these transcripts, in most cases Eugene does not actually reveal the basis of his character. However if a machine is to imitate human conversation then having a character of some sort is important otherwise judges would more easily discover the machine's identity. In this case Eugene does throw in odd pieces of information because of his character, when it seems appropriate.

The right identification

The concept of what is and what is not a 'right identification' is very important as far as the 30% pass mark is concerned and we have taken a relatively strict approach in this sense. For example one viewpoint is that for a judge to make the 'right identification' they must correctly identify both the machine as being a machine and the hidden human as being a human (Traiger, 2000). This means that any other decision on the part of a judge would not be a 'right identification', thereby including cases in which either the machine is selected as a human and/or a human is selected as a machine. Also included are cases in which the judge is unsure about either of both entities.

What we have included here however are only the cases in which the machine Eugene Goostman was itself not correctly identified, either being referred to as a human or where the judge was unsure about it. As can be seen from the previous section, the judges then made different decisions concerning the human in parallel in each case, sometimes selecting it as a machine, sometimes as a human and sometimes being unsure. What we have not included however are cases in which Eugene was selected as being a machine but where the parallel human in each case was incorrectly selected as being a machine and/or the judge gave an unsure mark against the human.

With our strict grading procedure, 10 judges (out of 30) marked Eugene incorrectly (i.e. in 10 cases they did not make the right identification), giving a final judge score of 33% – thereby surpassing Turing's 30% mark. These are the 10 cases that relate to the transcripts presented in the article.

If it is considered, however, that incorrect identification of the human also counts to the total, even where the machine has been correctly identified, then there were a further three cases which occurred in these experiments. Both Judge 9 and Judge 14 in fact decided that Eugene was a machine but then also came to the incorrect conclusion that the hidden human was a machine. For Judge 11, having decided that Eugene was a machine, they were unsure about the nature of the hidden human entity. Meanwhile the case of Judge 16 is also rather curious. While they marked Eugene as being a machine, they then scored it 100 out of 100 for humanlike conversation.

If we also take the marks of judges 9, 11 and 14 into account (as well as the 10 judges in the transcripts presented) then in the experiments a total of 13 judges (out of 30) did not make the right identification – in terms of both correctly identifying the machine and the human, giving a final judge score of 43%. We need to point out here however that the authors prefer the stricter line that we have taken, meaning that the ‘right identification’, as stated by Turing, we understand to refer to the machine only. Hence we believe the 33% to be the more appropriate grade.

For statistical purposes we have merely looked at the identification of hidden entities in terms of whether they were a human or a machine. If a judge decided that an entity was a human then they were asked for further details, for example, was the human male or female. If this gender classification was incorrect on the part of a judge nevertheless the identification was still regarded as being correct because a (human) entity had been correctly classified as a human. However such gender blurring issues will be considered in other articles.

Machines and judges

There were five machines involved in total in these tests and their success rates were as follows: Eugene Goostman 33%, Elbot 27%, JFred 20%, Ultra Hal 13%, Cleverbot 7%. In each case their success rate is in respect of judges not identifying them correctly as being machines. Results for the same machines from the 2012 Bletchley Park tests were discussed previously in Warwick and Shah (2014a, 2014c).

A wide variety of judges and hidden humans were involved in the tests, young and old, male and female, etc., in that we were attempting to aim towards Turing’s statement concerning ‘average interrogators’ (Turing, 1950). The ploys and strategies of the machines were applicable to all judges; however, machines did not exhibit different strategies for different judges, other than those that came out in the conversation. There were some gender issues and these are part of our on-going research.

What we have looked at in this article specifically are cases in which one machine was not correctly identified, the converse case of humans being misidentified as machines is also of interest, although for very different reasons. Such cases have been looked at (in terms of previous tests) in depth elsewhere (Warwick et al., 2013; Warwick & Shah, 2015). It must be remembered that humans are all very different and some can exhibit such features as to be spontaneous, draw interesting relationships, provoke, try to control the discourse and use language, examples and knowledge that a judge may not understand. All these features are likely to assist in a human being misidentified as a machine and are completely different characteristics to those we have looked at in this article, where machines are, arguably, attempting to do quite the opposite.

It must be remembered that an important feature of the Turing test is not whether a machine gives a correct or incorrect response or a truthful or untruthful one, but rather if it gives the sort of response that a human might give, such that an interrogator cannot tell the difference

(Warwick, 2011) between the machine and a parallel human foil. One ploy that we witnessed here on several occasions by Eugene was that of not directly answering a question but rather attempting to steer the conversation either by firing a question back to the judge or even by changing the subject (Warwick, 2012). This is a technique that often works well in everyday human life and clearly has had, it appears, a dramatically positive effect here. But it is just the sort of thing that humans do – indeed politicians are usually very good at this tactic in interviews.

One interesting aspect that did not seem to appear in these transcripts is the previously witnessed ploy by some judges to ask factual questions for which they expect the entity to have a certain knowledge base and therefore a particular type of answer is expected. In particular it has previously been shown (Warwick & Shah, 2014b) that interrogators sometimes exhibit prejudiced assumptions as to what they believe should be widely known and subsequently conclusions are drawn as to the nature of an entity based on whether it does or does not appear to know a particular fact known to the interrogator.

To adhere strictly to Turing's wording we have focused here, as best we could, on 'average interrogators' (Copeland & Proudfoot, 2008; Turing, 1950) and have included a wide variety of people. It will be interesting subsequently to see how such machines perform when taken to task by experts only. In terms of Eugene's success, we can point to a number of tactics that have helped its success: 1. It has a character, 2. It frequently poses pretty stock questions to the interrogator, 3. It occasionally throws in a spelling error. Combined with this the occasional use of humour on the part of hidden humans seems to help the machine cause. The average length of human conversations was substantially lower than the equivalent measure for machine conversations.

However it must be said that in this article we have put forward a collection of machine discourses and no matter what we think of the quality of those discourses ourselves, the human judge in each case was not able to identify the entity as being a machine. Implicit in that conclusion one can argue that, based on Turing's statements and the mark obtained, is that the hidden entity thinks. On the other hand, it could be said that such a decision has merely been made based on a brief, 5 min, conversation and it could be argued to be no more than if it looks like a duck and quacks like a duck then it is a duck (French, 2007). If nothing more, the test certainly fuels the philosophical argument.

The Turing test is an important benchmark in terms of one aspect of AI and its philosophy but it also paints an important picture of the trusting and relatively reactive way in which humans communicate, thereby highlighting a richness of discourse as well as some of its inadequacies.

Conclusions

In this article we have selected 10 specific transcripts arising from 10 of the 30 judges who took part in the June 2014 experiment. It should not be considered that the remaining 20 judges were perfect. On several occasions, not just those already pointed out, a judge decided that a hidden human was a machine, at other times a judge was simply unsure about a discourse and at times a different machine was considered to be human. In some cases it is obvious to see why such a selection was made, however, in other cases we have had considerable difficulty in understanding why a judge made the decision they did. What we have presented here are cases when a specific machine was either considered to be human or where the judge was not sure that it was indeed a machine that they were conversing with and in each case we have attempted to unravel some of the thought processes involved.

One important aspect of the 2014 test results is that a machine has now achieved a 33% deception rate, which is the first time in our studies that the 30% barrier has been broken. Indeed there has been a steady improvement in the best machine's performance from 2008, through 2012 and now to 2014. It is evident from all of these that experimentation involving practical Turing tests certainly gives us an indication of the standing of present-day machine conversation systems, as was looked at, in terms of the 2012 tests, in Warwick and Shah (2014c). However it also gives us more of an insight into how humans communicate, how they can be affected by normal conversational points such as lies and jokes and how they can be fooled into believing something that is some distance from the truth.

The 10 Eugene Goostman transcripts discussed in this article, along with others from the 2014 tests in which hidden humans were incorrectly classified as a machine or an unsure classification was returned, are now being further processed for post-Turing test analysis by a different set of independent judges. Without telling a second set of judges how the Royal Society test judges scored, we wish to learn if the former set decide the same way or differently from the latter set of judges – another aim of the second experiment is to find if there is a particular group more susceptible to deception, a crucial factor to improving trust in cyberspace transactions.

Acknowledgements

The authors would like to thank the Royal Society for hosting these tests and those who took part as judges and reviewers.

Disclosure statement

No potential conflict of interest was reported by the authors.

Developers

We thank the developers for accepting our invitation allowing their machines to participate in this experiment: Rollo Carpenter and Existor for *Cleverbot*, Fred Roberts and AI Solutions for *Elbot*, Robert Medeksza and Zabaware for *Ultra Hal*, Robby Garner and the Turing Hub for *JFred*, and last, the winner of the best machine *Eugene Goostman* from Vladimir Veselov and his team.

Note

1. Email: h.shah@coventry.ac.uk

References

- Copeland, B. J. (Ed.). (2004). *The essential Turing – The ideas that gave birth to the computer age*. Oxford: Clarendon Press.
- Copeland, B. J., & Proudfoot, D. (2008). Turing's test: A philosophical and historical guide. In R. Epstein, G. Roberts, & G. Beber (Eds.), *Parsing the turing test: Philosophical and methodological issues in the quest for the thinking computer* (pp. 119–138). USA: Springer.
- French, R. (2000). Peeking behind the screen: The unsuspected power of the standard turing test. *Journal of Experimental and Theoretical Artificial Intelligence*, 12, 331–340. doi:10.1080/09528130050111464
- French, R. (2007). If it walks like a duck and quacks like a duck... The Turing test, intelligence and consciousness. In P. Wilken, T. Bayne, & A. Cleeremans (Eds.), *Oxford companion to consciousness*. Oxford: Oxford University Press.
- Harnad, S. (1992). The turing test is not a trick: Turing indistinguishability is a scientific criterion. *ACM SIGART Bulletin*, 3, 9–10. doi:10.1145/141420.141422

- Hayes, P., & Ford, K. (1995). Turing test considered harmful. *Proceedings of the International Joint Conference on Artificial Intelligence, Montreal (Vol 1, pp. 972–977)*.
- Hodges, A. (1992). *Alan Turing: The enigma*. New York, NY: Vintage Press.
- Moor, J. H. (2003). The status and future of the Turing test. In J. H. Moor (Ed.), *The Turing test—The Elusive standard of artificial intelligence* (pp. 197–214). Dordrecht: Kluwer.
- Shah, H., & Henry, O. (2005). Confederate effect in human-machine textual interaction. *Proceedings of 5th WSEAS International Conference on Information Science, Communications and Applications (WSEAS ISCA), Cancun, Mexico, May 11–14* (pp. 109–114).
- Shah, H., & Warwick, K. (2010a). Hidden interlocutor misidentification in practical Turing tests. *Minds and Machines, 20*, 441–454. doi:10.1007/s11023-010-9219-6
- Shah, H., & Warwick, K. (2010b). From the buzzing in Turing’s head to machine intelligence contests. In *Proceedings of symposium for 1st towards a comprehensive intelligence test. AISB Convention*. Leicester: De Montfort University.
- Shah, H., & Warwick, K. (2010c). Testing Turing’s five minutes, parallel-paired imitation game. *Kybernetes, 39*, 449–465. doi:10.1108/03684921011036178
- Shah, H., Warwick, K., Bland, I., Chapman, C. D., & Allen, M. J. (2012). Turing’s imitation game: Role of error-making in intelligent thought, Turing in Context II, Brussels. 10–12 October, pp. 31–32, 2012. <http://www.computing-conference.ugent.be/file/14> – presentation available here: http://www.academia.edu/1916866/Turings_Imitation_Game_Role_of_Error-making_in_Intelligent_Thought
- Traiger, S. (2000). Making the right identification in the Turing test. *Minds and Machines, 10*, 561–572. doi:10.1023/A:1011254505902
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind, LIX*, 433–460. doi:10.1093/mind/LIX.236.433
- Warwick, K. (2011). *Artificial intelligence: The basics*. London: Routledge.
- Warwick, K. (2012). Not another look at the Turing test! In M. Bielikova, G. Friedrich, G. Gottlob, S. Katzenbeisser, & G. Turan (Eds.), *Proceedings of SOFSEM 2012: Theory and practice of computer science, Lecture Notes in Computer Science, Vol. 7147* (pp. 130–140). Berlin Heidelberg: Springer.
- Warwick, K., & Shah, H. (2014a). Good machine performance in turing’s imitation game. *IEEE Transactions on Computational Intelligence and AI in Games, 6*, 289–299. doi:10.1109/TCIAIG.2013.2283538
- Warwick, K., & Shah, H. (2014b). Assumption of knowledge and the Chinese room in Turing test interrogation. *AI Communications, 27*, 275–283.
- Warwick, K., & Shah, H. (2014c). Effects of lying in practical turing tests. *AI and Society*. doi:10.1007/s00146-013-0534-3
- Warwick, K., & Shah, H. (2015). Human misidentification in Turing tests. *Journal of Experimental and Theoretical Artificial Intelligence, 27*, 123–135. doi:10.1080/0952813X.2014.921734
- Warwick, K., Shah, H., & Moor, J. H. (2013). Some implications of a sample of practical Turing tests. *Minds and Machines, 23*, 163–177. doi:10.1007/s11023-013-9301-y
- Whitby, B. (1996). The Turing test: AI’s biggest blind alley? In P. J. R. Millican & A. Clark (Eds.), *Machine and thought: The legacy of Alan Turing (Vol. 1, pp. 53–62)*. Oxford: Clarendon Press.