

The April Fool Turing Test

Mark S. Dougherty, Sofi Hemgren Dougherty and Jerker Westin

*Department of Culture, Media and Computer Science
Högskolan Dalarna
78188 Borlänge
Sweden*

{mdo, shd, jwe}@du.se

Abstract: This paper explores certain issues concerning the Turing test; non-termination, asymmetry and the need for a control experiment. A standard diagonalisation argument to show the non-computability of AI is extended to yields a so-called “April fool Turing test”, which bears some relationship to Wizard of Oz experiments and involves placing several experimental participants in a symmetrical paradox – the “April Fool Turing Test”. The fundamental question which is asked is whether escaping from this paradox is a sign of intelligence. An important ethical consideration with such an experiment is that in order to place humans in such a paradox it is necessary to

fool them. Results from an actual April Fool Turing Test experiment are reported. It is concluded that the results clearly illustrate some of the difficulties and paradoxes which surround the classical Turing Test.

Keywords: Turing Test, AI, Gödel's Incompleteness Theorem, Diagonalisation, Symmetrical Paradox.

Acknowledgement: The assistance of students and staff at an anonymous university are gratefully acknowledged (anonymity was a condition of participation in the experiment).

1 Introduction.

In a seminal paper Alan Turing set out his famous impersonation game, which soon became known as the Turing Test (TT) of artificial intelligence (AI) (Turing, 1950). Few papers touching on the field of computer science have fuelled such controversy. Some authors have hailed this paper as the birth of the study of artificial intelligence, whilst others have dismissed the test as irrelevant and badly designed. But as Oscar Wilde once remarked (Wilde, 1890):

Diversity of opinion about a work of art shows that the work is new, complex, and vital. When critics disagree the artist is in accord with himself.

Personal opinions aside, nobody could deny that Turing greatly influenced those who came after him. (Saygin, et al , 2000) gives a comprehensive review of the Turing Test and the debate it has inspired in the 50 years following the publication of Turing's original article. An important point which the authors mention in this paper is that Turing's original paper involves a somewhat obtuse gender aspect which makes his intentions slightly unclear. Most authors have chosen to ignore this additional complication and settled for a “standard format” of the test as follows:

The standard format of the test concerns three agents:

- A human interrogator
- A human respondent
- A machine (AI) respondent

The task of the interrogator is to determine which respondent is the human and which is the machine. To do this, the interrogator must hold a conversation with each of the two respondents. The machine “wins” (and is declared to exhibit intelligence) if a series of interrogators find it indistinguishable from the human respondents. A simple schematic is shown in figure 1:

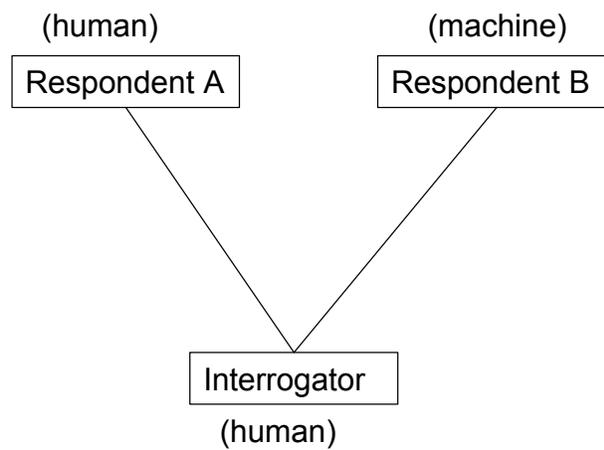


Figure 1: The “Standard” Turing Test

The communication required is achieved by remote chat messages, to prevent physical appearances immediately giving the game away. The communication link is represented as a simple line on the schematic. Some authorities have disputed the validity of this simplification and introduced the terminology “Total Turing Test” to mean a fully robotised version where the machine must display all of the physical appearances of a human being (Harnad, 1989).

2 Introduction.

It is interesting to note that Turing himself did not specify any particular time limit for the interrogation to take place within. He was of course aware of this problem and notes within his section “the mathematical objection” the following:

If it (the machine) is rigged up to give answers to questions as in the imitation game, there will be some questions to which it will either give a wrong answer, or fail to give an answer at all however much time is allowed for a reply.

Turing’s answer to the problem of the respondent not replying to a question is to argue that a human being could be fallible too. Thus he saw no particular problem in the possible non-termination of the interrogation process. However, the situation described is only a special case of a more general situation. Non-termination is also possible in the case where the conversation simply continues ad infinitum. This case is not discussed in Turing’s paper.

This oversight is strange. Obviously Turing would have been fully conversant with the issues surrounding termination through his earlier work in computability theory (Turing, 1937). Yet a relatively simple diagonal construction illustrates the importance of non-termination to the Turing Test. It also provides a much more intuitive insight to the Gödellian issues at stake. These have also been discussed at length (Lucas, 1961, 1996), but the treatment seems to be generally very abstract and hard to follow in much of the literature.

Something tantalisingly close to a diagonal construction is the so-called inverted Turing Test. (Watt, 1996) describes the idea behind the ITT as follows:

Instead of evaluating a system's ability to deceive people, we should test to see if a system ascribes intelligence to others in the same way that people do.....a system passes if it is itself unable to distinguish between two humans, or between a human and a machine that can pass the normal Turing test, but which can discriminate between a human and a machine that can be told apart by a normal Turing test with a human observer.

Again, no discussion of the problem of time-limiting the interrogation is presented. However, Watt does discuss the possibility of a machine interrogating a copy of itself, but rules this out, arguing that the machine would be likely to have privileged information.

Figure 2 shows the experimental set-up alluded to by Watt:

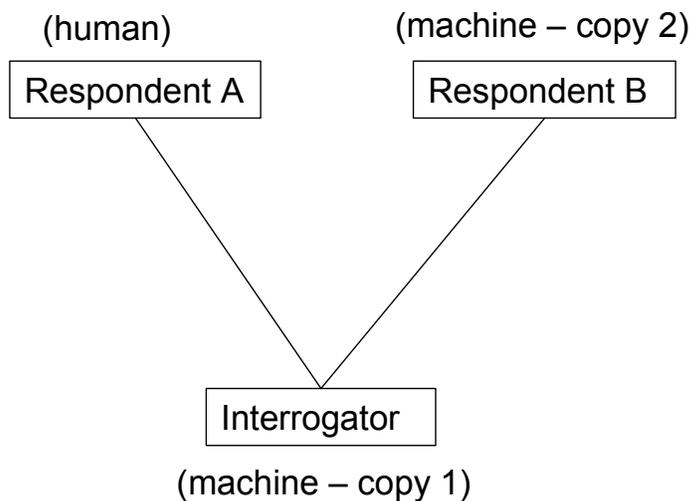


Figure 2: The Inverted Turing Test

Copy 1 of the machine is in the role of the interrogator, and is asked to distinguish between two respondents. One respondent (labelled A) is a human and the other respondent (labelled B) is copy 2 (of the same machine as the interrogator).

Let us first assume that the interrogator, after some finite period of interrogation gives an answer:

- “A is human and B is a machine”. This might seem like the right answer; copy 1 has passed the ITT. But at the same time it seems that copy 2 has failed the TT.

- “A is a machine and B is human”. Now copy 1 fails the ITT and copy 2 passes the TT.

So we have a paradox – the Turing Test is susceptible to diagonalisation arguments just like any other non-recursively enumerable problem in computability.

French (1996) declares the ITT to be of no interest, using the construction showed in figure 3:

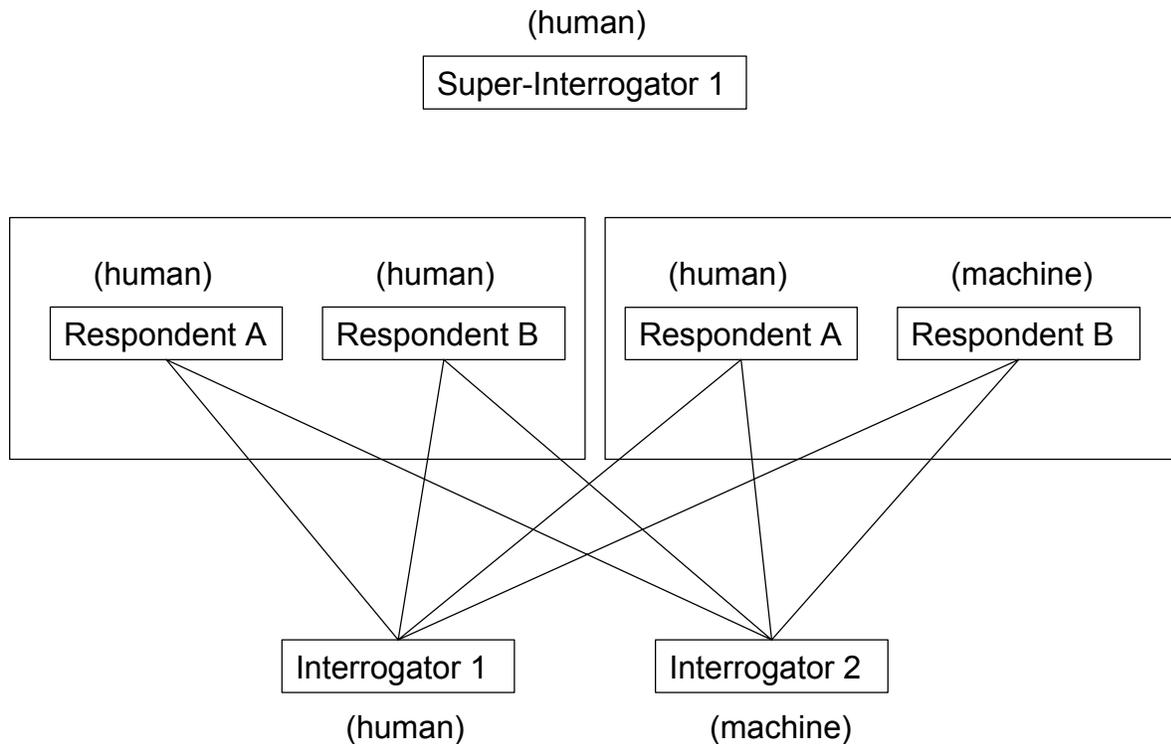


Figure 3: French's Reduction of the ITT to a Standard TT

The idea behind the construction is to reduce the ITT to a standard TT. Each interrogator is performing an ITT (which can be against two human respondents or one human and one machine). The super interrogator is observing the process and making a judgement about which interrogator is human and which is a machine.

French's construction is ingenious, but I feel that he has somehow missed the full implications of Watt's paper, which is the possibility of a diagonal construction. This in itself doesn't tell us anything we didn't already know, but it can lead us in a new direction.

3 April Fool Turing Tests.

How can we escape from the diagonal paradox which the ITT exposes us to? One could argue, as Watt does, that privileged information makes this test biased. I don't buy this argument. Assuming that we get an answer, the machine will have cooked its own goose and the question of how or why this happened doesn't seem very relevant.

A second escape path is that the interrogation process will continue indefinitely, i.e. it will fail to terminate. Obviously the human in the game would eventually get too tired, but this can be dealt with by replacing the human with a copy 3 of the machine (figure 4):

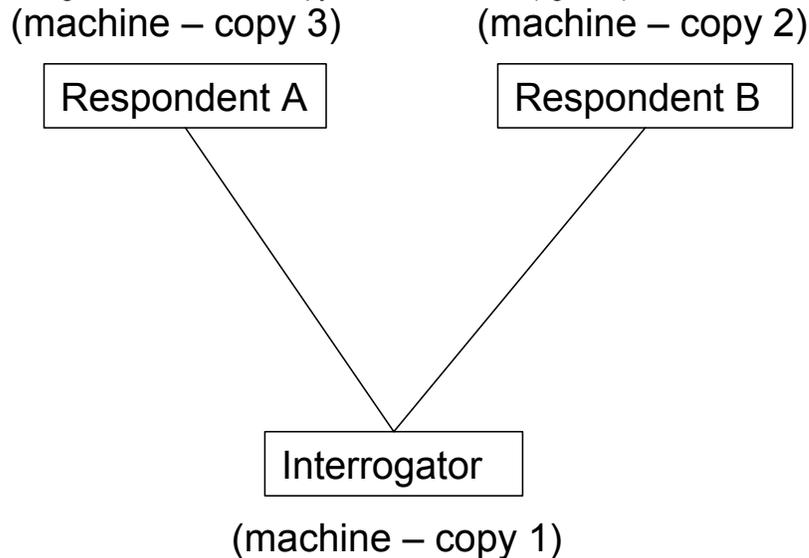


Figure 4: April Fool Turing Test

An interesting point to note about this experimental set-up is that we must now lie to copy 1, the interrogator. This is because we are asking it to distinguish between a human and a machine, when in fact we are fooling the machine because both A and B are now machines. The term “April Fool Turing Test” seems appropriate for this kind of set-up. Note that a similar paradox still emerges if the interrogator returns an answer:

“A is a machine and B is a human” is simultaneously declaring itself to be both machine and human at the same time.

Can this set-up really compute forever without returning an answer? The answer to this must be in the affirmative. There is no restriction on the size of message transmitted and the system is non-Markovian – its “state” is dependent on the entire history of the conversation held between the parties.

A third escape path is that the interrogator (copy 1) realises the paradoxical nature of the situation it has placed in, i.e. it escapes out of the Gödelian loop it is ensnared in by announcing a refusal to play any more. But this seems impossible, because it is opposed by one or more copies of itself. The copies can anticipate any self-awareness knowledge that copy 1 has. Of course copy 1 can anticipate this reasoning and so on, but there is no escape from what is obviously an infinite regress.

Part of this difficulty seems to stem from the fact that copy 1 and copy 2 are acting in different roles in a directly adversarial fashion. This is obviously the foundation upon which the diagonal argument is built. We can eliminate this by incorporating a further communication link (figure 5).

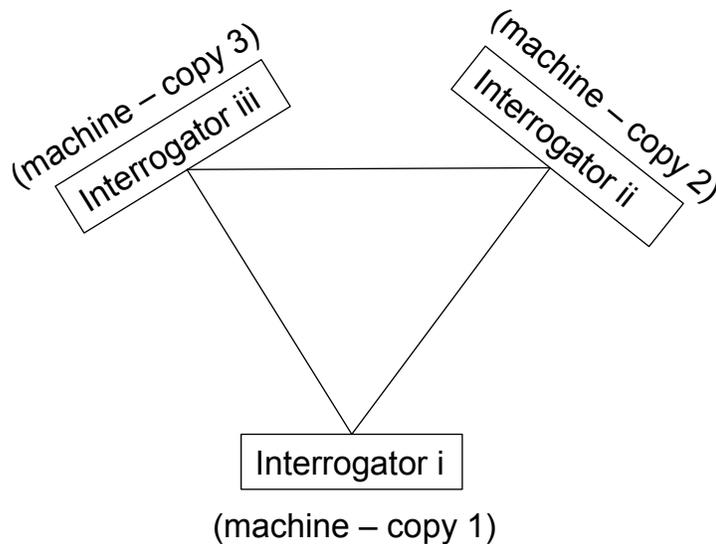


Figure 5: Symmetrical April Fool Turing Test

If all three copies are asked to communicate with their two opponents and determine which is human and which is a machine, this results in a Symmetrical April Fool Turing Test. All three machines are now simultaneously acting the part of interrogator (knowingly) and opponent (unknowingly). Note that all three participants are now labelled as interrogator (i, ii, iii respectively). A key difference now is that we (the humans who devised the experiment) are fooling the machines, but the machines are not deliberately trying to fool each other.

Again, if any of the machines return a definite answer we fall directly into a paradox. Similarly, the conversation could go on forever. Thus far nothing has changed. However, the question as to whether the machines could now escape from this modified Gödelian paradox seems interesting. We do not seem to have the same situation as before. Several escape routes are possible:

- a) One of the copies realises it is faced with only machine respondents (unlike what it has been told) and declares a refusal to play any more.
- b) One or more of the copies recognises that one of its respondents is a copy of itself and declares a refusal to play any more.

Both of these escape routes represent reasonably intelligent actions, although Watt's argument about bias and privileged knowledge still applies to escape route (b). Neither seems vulnerable to Gödelian arguments, since the machines are not acting in a directly adversarial fashion. So an exit from the experiment along the lines of options a and b could be seen as evidence of machine intelligence. Does a diagonal paradox still exist? The answer to this is in the affirmative, although it requires some rather contrived thinking to imagine one, described as option (c) below:

- c) One or more of the copies realise that they have been fooled by the human experimenter and deliberately contrive to sabotage the experiment without the experimenter knowing (for

example by continuing the conversation ad infinitum or lying to the experimenter about their reasons for exiting).

For the experiment to work, an element of surprise also seems essential. The main problem with April fool jokes is that they are always played on April 1st – when I read the newspaper on this date I am always on the look out for ridiculous spoof articles. If the machine is aware that it will be faced with this situation (and given that it, or at least its developer, might have read the contents of this article!), then the experiment could fail.

In short, the Gödellian loop is now between experimenter and experiment rather than within the experiment itself. A parallel can be drawn with French's "super interrogator" standing outside of the ITT. The terrible conflict between creator and created depicted in Mary Shelley's Frankenstein also springs to mind.

One further option is to play the symmetrical version of the April Fool Turing Test with different machines, instead of several copies of the same machine.

4 A Proposed Experiment.

The April Fool Turing Test has another interesting aspect. We can invert it and play it with human subjects (figure 6).

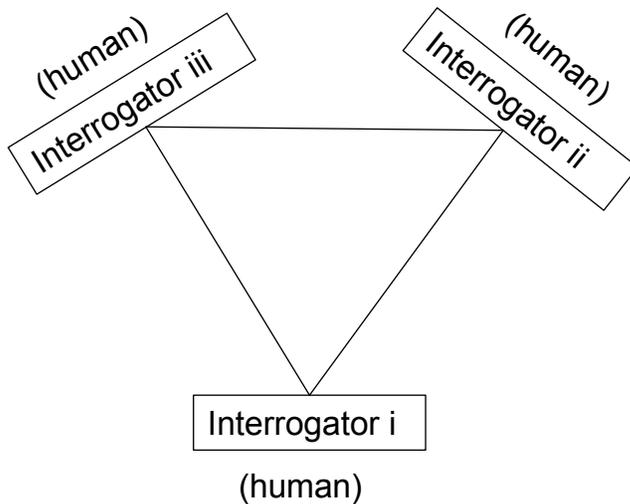


Figure 6: Symmetrical April Fool Turing Test with Human Interrogators

In the experiment, the three human subjects are each told they are to act as the interrogator in a normal Turing Test. In fact they are playing against each other in an April Fool Turing Test. Exactly the same outcomes are possible as for the machines. Just as when the machines are playing, we, the experimenter, can be fooled by the participants just as we are trying to fool them.

We could of course do a similar experiment with a non-symmetrical April Fool Turing Test, i.e. only the interrogator is being fooled. See figure 4, but replace the three machines with humans.

This experiment would however suffer from some of the disadvantages alluded to earlier – the participants are not on an equal footing and this could bias the outcome.

The question we wish to have answered by carrying out such an experiment is whether humans can escape from the paradoxical loop they are placed in. Interviews with the participants afterwards will help to elucidate how such an escape takes place. But, in the final event the experimenter is still placed in a Gödelian loop; the participants, realising that they have been fooled, can in turn lie to us. With human subjects it might be possible to control for this (using a polygraph lie detector test) but this raises some difficult ethical issues and is not considered further.

Even this experiment raises an important ethical issue. For the experiment to work, the participants have to be deceived by the experimenter. Is this justifiable? We answer in the affirmative. Firstly it seems highly unlikely that anybody could be harmed by participating. On the contrary, taking part could actually be an amusing and interesting experience. Secondly, “April Fool” jokes are in general not seen as taboo in western society. “White lies” are also accepted as reasonable behaviour by most people. Finally, it is of course possible to explain the reason for the deception after the experiment takes place.

An interesting aspect of this experiment is that it provides a control to the Turing test. The Turing test assumes that people are good at distinguishing between a machine and a human. Testing this point is difficult, because we don't (yet) have very convincing AI. This test asks whether people can accurately determine that a supposed machine is in fact human. Because it uses no AI machines, we can carry out such a test. If humans pass this test it suggests that they are rather good at making the distinction between human and machine (although of course their attitudes will inevitably be coloured by their upbringing and social norms).

5 Experimental Set-up.

The proposed experiment was performed with a number of student volunteers in the computer science department of a university. Four groups of three participants were placed in the triangular paradox described. The participants were simply told that they were asked to perform a Turing Test (this was briefly described to make sure they knew what it was). Each student was in a separate office, with communication arranged through two MSN messenger windows, each under a pseudonym. The experiment lasted for 20 minutes at which point the participants were interrupted. The volunteers were informed that they were free to terminate the experiment at any time and were asked not to use rude language or be otherwise offensive. They were also told that MSN had been set up to record the conversations. As is usually the case with Turing Tests, the subject matter was restricted to a particular domain. In this case the subject matter was set to “April Fool's Day”.

It was not so easy to start the experiment. It was of course necessary that all three subjects started conversing at the same moment and this required some careful handling of the logistics in order to achieve a synchronised start. However, conversations in all four groups were successfully initiated.

When the experiment was completed, each of the participants was offered to take part in an individual interview immediately after. All twelve students who participated in the experiment agreed to be interviewed. The shortest interview took around 15 minutes and the longest interview took around 30 minutes.

The interviews after the experiment were carried out to gain an insight into:

- a) the participants' ideas during, and about, the experiment,
- b) the participants' general views on AI as well as on intelligence testing in general
- c) whether the participants were upset by learning the truth (and also to give them a chance to communicate on a one-to-one basis if they felt that the experiment was unethical in some way)

The interviews were carried out in a semi-structured way with one person being interviewed at a time. A mini-disc was used to record the interviews (all participants agreed to this).

The choice of a semi-structured interview was selected to:

- a) be able to have a set range of topics covering the issues we wanted to gain an insight into (this would enable us to compare the answers and to get an overview of what was preferred/chosen by the participants in general)
- b) be able to have open ended questions and to provide a degree of freedom for the participants (to be able to focus upon what he/she felt was specifically important within the range of topics covered).

Participants were informed both before the experiment and before the actual interview that they were free not to take part in the interview. They were also informed that they did not have to answer a question they did not want to answer and that they were free to cancel the interview at any stage during the interview, or after the interview. They were further informed that their names would not be used in any reports, nor would the place where the experiment and the interview took place be given an exact location in published articles.

6 Results.

The groups of students who were interviewed have been given false names. It is not even possible to infer the gender of the participant from the names. We have chosen to give 3 of the participants female names and 9 male names. The reality was that 2 female students participated and 10 male students. We have however kept the students in the same groups as they took part in for real. It is therefore possible to compare the students' ideas and thoughts within each group. The following groups of student were interviewed (note that the order of presentation is random and does not relate to the real situation).

Group 1: Eric, Emma, Thomas

Group 2: Clara, Hugh, Jim

Group 3: Anna, Colin, Ben

Group 4: Carl, John, David

When one of the students, "Colin", was interviewed the recorder stopped working midway through the interview. Colin agreed to come back the next day for an extra interview where we covered all the areas again. He was the one student who had been most surprised about the actual reality of the experiment (and did in fact not believe the real set-up of the experiment when presented with the facts during the first interview). In the material below Colin's views have been taken into account regardless of whether they came from the initial interview or the follow-up interview.

The results from the interviews have been broken down and analysed under four categories:

- Part one: Prior knowledge and incentive to participate
- Part two: Thoughts/particular moments during the experiment
- Part three: Thoughts in retrospect about the experiment
- Part four: Qualitative interviewing

In the interviews further topics related in various ways to the area were also covered. The data gained from these parts of the interviews goes beyond the scope of this paper and is hoped to be added to further studies and presented in a future scientific paper.

Appendix one contains the manual used during the semi-structured interviews.

Appendix two contains selected extracts from the transcriptions of the interviews. This material concentrates on parts two and three (as defined above).

6.1 Part one: Prior Knowledge and Incentive to Participate.

6.1.1. Why Participate?

Why did students chose to participate in the experiment? Three of the students, Ben, Hugh and Jim have chosen to come because of their particular interest and curiosity of the field. Jim in particular describes how he was very intrigued about being able to meet and challenge a real AI. Hugh is also keen to test a real AI “and see whether it really works or not”. Ben simply states that he thought the experiment sounded interesting. Six of the students, Anna, Carl, Clara, Colin, David and John all refers to various reasons based around that they want to help out, either in general for research or because a teacher suggested this experiment to them. Three of the students do not really have any particular answer to give why they chose to participate. One of them just laughs happily and states: “...don’t know – to get a t-shirt!...” (Eric). These results are summarised in table 1:

A particular interest in AI or the TT	To support research in general (or happy to “help out” when a teacher asks...!)	Don’t know... (“...to get a t-shirt”)
4	5	3
Ben, Colin, Hugh, Jim	Anna, Carl, Clara, David, John	Emma, Eric, Thomas

Table 1: Incentives to Participate in the Experiment

6.1.2. Prior Knowledge of AI/TT.

Four of the students participating had studied AI and knew particularly of the TT and what to expect and not to expect of a TT. Four of the students claim that they only have some knowledge of AI. Four of the students expressed that they had not much knowledge of AI as they were first year students. The students who participated were all computer science students at the same university but were in different years.

None	Some knowledge	Studied AI/TT
4	4	4
Anna, Emma, Eric, Thomas	Ben, Carl, Clara, John	Colin, David, Hugh, Jim

Table 2: Prior Knowledge of AI/TT

6.1.3. Experience of human intelligence tests.

Most of the participants had tried an intelligence test for humans. One of the participants, Jim, did his first real intelligence test age five year old. Most participants referred to tests found on the internet. Three students claimed never to have done an IQ test of any sort; Anna, Ben and John.

Not all of the students chose to discuss this area. Clara, Eric, Jim and John points out that IQ tests work in some ways and measure at least some part of human intelligence. Jim criticises the fact that it is so easy to score high on the IQ tests if you are into maths. Clara points to the same fact and criticises how easy it is to spot a pattern in the tests. Anna, Ben, Carl, Colin and Hugh are very negative regarding human IQ tests. Carl points out that you can test a person seven days a week and still get different results. Hugh points out that everybody has different intelligence and that it cannot be measured. Colin refers to Gardner's theory on multiple intelligences, which he firmly believes in.

6.1.4. Beliefs Regarding the Existence of Real AI

The students were all asked what they would find best when measuring AI: relative or absolute quantities. One students preferred an absolute way of testing for AI; Hugh. He argues that it would be too complex to measure in a relative way and states "that many people would be afraid of it". He therefore suggests an absolute way of measuring which he finds the absolutely easiest and the most interesting. Nine other students however prefer having a relative scale instead of the absolute. Anna, Ben, Carl, Colin, David, Eric, Jim, John and Thomas all agree here. According to Jim absolute would not make any sense. He does not understand the point of saying that something is absolutely AI even if something is built to do a job and does the job. According to him there is no absolute AI but a relative scale makes sense. Clara and Emma do not express any particular views on this topic.

6.2 Part two: Thoughts and Particular Moments During the Experiment

In group 1 all of the participants decided that one of the respondents is a computer and one is a human. Eric particularly focuses his decisions around that one of the respondents, the one he believes is a computer, did not for example remember earlier conversation and also appeared "stiff". Emma expresses that it was fun to take part in the experiment and bases her decisions on why one of the respondents is a computer on that it forgot that it was April Fools day tomorrow as well as the other respondent being able to always provide answers. Thomas points out that when one of the respondents knew the characteristics of a teacher being discussed then he was certain that this was a human respondent. He therefore concluded that the other respondent must have been a computer.

In group 2 Clara found the experiment boring and the topic difficult to discuss around. She was also concerned because of the difficulties she had experienced with writing in English. She was convinced that she had found a respondent which was a computer as this one was both more formal than the other and also came up with answers very slowly, which she believed to be a programmed delay function. Hugh is very sure about which is a computer and which is a human and says that it was not a difficult task. He had asked a question and then another question. The respondent he believed to be a computer only answered the last question every time. He felt strongly that this respondent had a typical machine behaviour. He had initially believed that a computer would reply quickly but realised that this was not the case here. The conversation he had with the other respondent on the other hand he describes as interactive and more alive (my word). Jim is very unsure and very concerned about the outcome of the experiment. He has been trying to challenge the respondents by trying to take the conversation to a higher level but he is not sure it worked. He has not been able to formulate an answer and he has not been able to decide which is a computer and which is a human. In the end he decides that the respondent with short answers is a computer whereas the other one with more complex answers is a human.

In group 3 Anna found the experiment really interesting but the topic difficult. She point out that she initially found it strange to use MSN for the experiment but that she soon forgot about this. She felt she knew which was a computer when she got an odd answer regarding a practical joke she had discussed. She also noticed spelling mistakes by the respondent she believed to be a computer. Colin has studied AI and is particularly interested in TT. He spends the entire experiment actively challenging the other respondents and analysing their replies. He takes notes and compares and makes careful judgments. After about half of the time of the experiment has passed he is certain that he is communicating with two computers rather than with a computer and a human respondent. He bases his strong belief upon various facts such as noticed pattern in spelling mistakes and the delay but above all from the lack of empathy (our word) shown when discussing the pranks. All fits in. Ben decides to go open-minded into the experiment. He soon recognises his friend's sense of humour and feels that he knows which is a computer and which is a human.

In group 4 Carl notices that one respondent is using capital letters and end sentences with a period. She takes this as very formal and deduces that this is probably the respondent which is a computer. When she asks the question "Where are you from" the experiment quickly finishes as she soon realises who she is communicating with. She decides that the other one must be a computer. John receives only six responses and spends the short time the experiment is going on waiting for replies. He nevertheless forms some ideas which is a computer and which is a human. David has experience of "talking" with robots on computers and immediately recognises the pattern of having to wait for a long time for responses etc in the two respondents way of responding. He is certain that both of the respondents are computers until suddenly one of the respondents begin discussing where the participants origin from. Now he suddenly realises that he is communicating with his friends! He tries to fool them but the interview is terminated.

All of the students believed genuinely that they were doing a traditional TT where they would be talking to one computer and one person in spite of the fact that they were never informed that they were to take part in a TT but an "April Fools Turing Test". They were however expecting a traditional TT and did not enter the experiment expecting otherwise than to communicate with one human as well as with one computer. Only a few of the students realised that the set-up was not what they had expected. None realised without being probed that there were only humans involved (see table 3).

Computer + human (A)	Computer + computer (B)	Human + human (C)	Unable to decide
9	1 (-4)	0 (-11)	2
Anna, Ben, Carl, Clara, Emma, Eric, Hugh, (Jim), John	Colin! (very firm!) David (very firm to begin with – realises C later) Clara & Thomas (when asked if option B was possible! Chooses C in the end)	None initially. ALL except Colin, when provided with the options.	Thomas Jim very, very unsure

Table 3: Summary of the Judgements Made by the Participants

6.3 Part three: Thoughts in Retrospect About the Experiment.

6.3.1. New Ideas About Who the Respondents Were (Humans or Computers).

After the interviewee had described the experiment and given his or her view about what either of the respondents were, human or computer, the real situation was gradually revealed. The student was asked if he/she has thought about that we do not call this experiment simply a Turing Test, but an “April Fools Turing Test”. Many of the students quickly jump to various conclusions at this statement. Others just look surprised and confused at this. For those who are not triggered to make new guesses based on this I add: “If I tell you that you were not in fact communication with one human and one computer, as in a traditional Turing Test, but either with two people or two computers – what would your judgement be?”

In group 1 Eric felt that both respondents “tried to be” an AI and quickly guesses that both were humans after all. Emma also quickly guesses that her two respondents were humans but she still admits that one of them in her view could have been a computer. Thomas is unsure and cannot make up his mind; says that both could be both computers or both humans...

In group 2 Clara asks if both respondents were computers. When presented with the option that they could have been two humans she quickly responds that this is what she believes. Hugh has been very convinced about which is a computer and which is a human out of the respondents. When presented with the options he points out that there is no possibility that both were computers but says he could have guessed both were humans! Jim who has studied AI has been very unsure and concerned and found it very difficult to make up his mind about who was the computer and who the human. He is extremely relieved that both were humans. He had been seriously concerned when believing that as intelligent AI would exist as to disguise themselves well enough for him not to be able to judge.

In group 3 Anna finds it possible to believe that both could have been computers when asked but finds it most likely that both were humans. Colin has spent the whole experiment actively trying to challenge and analyse the respondents and is so absolutely convinced that both are computers that he even doubts my information when I reveal the true nature of the experiment. Ben on the other hand has recognised his friends sense of humour and can accept that one of the respondents may be a computer but no way both.

In group 4 the participant had a shorter session than the others due to the fact that Carl realised who one of the respondents was and terminated the experiment early. His firm assumption is that one respondent is a computer and the other one is a human and finds it strange with my questions. John finds it easy to buy that something was not what he had expected. He did not interact very much during the experiment partly due to the fact that the interaction time was slow and partly due to the fact that the experiment was finished early. David has initially been very sure that both the respondents were computers in a similar argument as Colin in group 3. He has, however, suddenly realised the true reality when the respondents ask and answer where they are from. He then decides to try to fool the others but the experiment is terminated before he has a chance to do this. None of the participants in this group was given the question “if I tell you...” but only asked if the name of the experiment would trigger new thoughts when reflecting over the experiment.

6.3.2. General Views Regarding the Fact that Participants could not be Fully Informed Beforehand.

Carl, Colin, David, Jim and John expresses firmly that they believe that this was necessary to be able to carry out this particular experiment. Anna, Emma, Eric, Thomas claim that it is Ok. Ben is slightly concerned. Clara and Hugh have no particular comments. Neither of the students expresses concerns regarding that participants could feel harmed in any real or lasting way by participating in the experiment.

In summary, we can state that the necessary lack of full disclosure about the experiment was not harmful in any way to the participants.

6.4 Part four: General Experience of Qualitative Interviews.

Only two participants, Ben and Jim, had any previous experience from participating in a qualitative interview. The rest had none.

7 Conclusions.

A simple truth is that any attempt to devise a perfect test for intelligence suffers from a severe difficulty, because we can never be certain of the motives and intelligence of those we are testing. If they are clever enough, they can fool us. As explained above, the most insidious Gödelian loop is between the experimenter and the experiment and there seems no way to avoid this.

However, the pessimism of the Gödelian argument should not stop us looking for tests which can be useful in practice. In this context the basic concept of the Symmetrical April Fool Turing Test seems to be of interest. It tests the basic ability of being able to “step outside of the box”. As is evident from the results presented in this paper, not even humans find this easy to do. The fact that some people found it hard to make a choice between two humans and two computers or even refused to believe the truth at all is extremely interesting.

In short, the Turing Test defines human intelligence in terms of a judgement made by human intelligence. This definition is quite obviously circular and it is not perhaps so surprising that it can lead to paradoxes and confusion.

8 References.

- Harnad, S. (1989), Minds, Machines and Searle, *Journal of Experimental and Theoretical Artificial Intelligence* 1(1), pp 5-25.
Lucas, J. (1961), Minds machines and Gödel, *Philosophy* 36, pp 112-127.
Saygin, A.P., Cicekli, I., and Akman, V. (2000), Turing Test: 50 years later. *Minds and Machines* 10:pp463-518.
Turing, A. (1950), Computing machinery and intelligence, *Mind* 59(236), pp 433-460.
Turing, A. (1937), On Computable Numbers, with an Application to the Entscheidungsproblem. *Proc. London Math. Soc. Ser. 2* 42, 230-265.
Watt, S. Naïve Psychology and the Inverted Turing Test. *Psychology*: 7, #14. <http://psycprints.ecs.soton.ac.uk/archive/00000506/>
Wilde, O. (1890). *The Picture of Dorian Grey*. Serialised in Lippincott's Monthly Magazine. Published by Wordsworth Editions in 1992.

9 Appendix One - Semi-structured interview manual.

Part one: Prior knowledge and incentive to participate

- What prior knowledge about a) general intelligence tests, b) Turing tests and other intelligence artificial testing, b) this experiment
- Incentive to participate?

Part two: Thoughts/particular moments during the experiment

- Describe the experiment in detail - as if I knew nothing about this experiment (how you went about it).
- What were your thoughts while you were doing the test

- Do you remember any particular moments
- What was it like to take part in this experiment (boring/fun?)

Part three: Thoughts in retrospect about the experiment

- What was tested in this particular experiment?
- Based on your conclusions from the experiment were you able to make a judgement about the respondents?

If the student realised the real situation:

- What made you realise what the real situation was?
- What went through your head when you saw this?
- Do you remember any particular moments?

If the student had *not* realised the real situation:

- If I tell you that you were not in fact communicating with one human and one computer, as in a traditional Turing Test, but either with two people or two computers – what would your judgement be?
 - What are your feelings regarding that participants could not be told the full details of the experiment in advance?
 - What does it feel like to be told afterwards/now?
 - Do you think others will be able to realise the real set-up?
 - o how long/short time do you think is needed?
 - o what do you think will make it possible for people to realise the true set-up?
 - o how do you think people will feel [when they come to realise that they have been placed in a paradoxical situation?] (surprise/confusion/anger)
 - What are your thoughts in general about tests where you are unable to have fully informed participants?

Part four: Qualitative interviewing

- Have you taken part in a qualitative interview (unstructured or semi-structured) before?

10 Appendix Two - Selected Quotations from the Interviews

10.1 Part two: Thoughts and Particular Moments During the Experiment

Group 1

- Eric
 - “If I was to guess I would say that XX was the computer and YY was the person”
 - “Just...XX didn't remember what I'd said before, kind of, it was a stiff conversation...the other one, YY, was a bit livelier”
 - “At first I tried to keep up the conversation. Then I tried to ask questions about things that had been brought up in the conversation before.”
 - “Quite a delay though...but pretty even between the both...”
- Emma

“The time went quickly. It was FUN, trying to analyse it and try to find out...”

“Particular moments: when it forgot! The one person forgot that it was April Fools day – so I thought: on the top of my head! This is probably a computer! And also, the other respondent always had some answers....so, yeah, I think I know.”

- Thomas

“We got the paper about what to do...(describes)... and then we went to the room and either one would be an AI and the other one would be a person...and I don't know...just started chatting...yeah, it was an easy topic...”

“There was a moment sticking that sticks out” [when the other person knew about the characteristics of a person mentioned]

“Regarding the other respondents...both gave subtle answers and...not sure...maybe...but...”

Group 2

- Clara

“The task was to...(describes)...and then I just started typing”

“Had to wait a long time – it was not so funny – boring...and the topic was not so easy...and particularly because my English is not so good and my spelling and so tried to use the words that I know – and then you maybe talked like a child! Yeah...and also this topic, you know, ...I don't know many words and how to spell...”

“When for example one of them get formal...and one of them had question marks...and the other more formal...Yeah, the one on the right...was it CC?...I thought it was a computer. Because it was more formal.”

“Also the TIME. The one that I thought was a computer was not that quick. But it could be a delusion...trying to delay...I don't know.”

- Hugh

“We started talking to two different persons, I mean respondents. I noticed one of them was quicker than the other...it was the same one who makes longer sentences...mmmm...”

“It was difficult to begin with. The point of doing the test at the beginning was just to talk...but after a few minutes to try to know if one was a computer...I think it was not too difficult. Since...one of them was always answering short answers and when you asked more (too many questions at the same time) he was answering like a machine.”

“I don't know how to describe it...you know you ask a question and you ask a second one. He just answered the last question – it is like a machine behaviour.”

“The first thing was the time of the answer. I was thinking the computer would go faster at the beginning...but the other one you could really feel you were really chatting with someone...like its interactive...you are TALKIN to someone...you know where you go, you know, you feel it, you have previous experience about DD...Yeah, I think it was (not at the beginning definitely but after a few minutes it was) easy.”

[What is your deduction?] “Oh, it WAS a machine! [very, very certain – laughs happily]

- Jim

Very unsure. Concerned nearly. States after some discussion that he believes: “the slower answer” = the computer, and “the complex answer” = the human.

Describes the test thoroughly. “sat down and begun speaking to both respondents and asking them questions, trying to find out...without going outside the given topic...yeah, and just trying to ask questions...”

“Pretty difficult. I was trying to get to a higher level to see if there was any kind of abstraction in the respondents answers or trying to use strange ways of thinking...and...I am not sure it worked because I am not sure which one was a computer...”

[So, you are not sure?]”...but I have an idea...but...mmm”

[What’s the idea based on...?] “Well obviously the answers, but it’s very difficult because I was assuming that the computer would probably try to trick me. Like one of the respondents was talking, writing in a very laid back language...it was clear that the respondent was not good at English or pretending not to be good at English and be able to spell correctly. The other one was giving more complex answers, and that might be some kind of trick also. It might be a computer just randomly generating so called elaborate thoughts without there being something

Real behind it...So I am not sure, but I think that the one giving the simpler answers was the computer”

“what happened was that one of the respondents, the one giving the complex answers, pasted the last part of a conversation back so as to, to...and I was wondering if...well obviously I don’t think a computer would have done this but on the other hand it is easy to programme a computer to have this feature...”

[So, you were sitting there wondering how well programmed the computer was?]

“YES! And I am really interested in this...and whatever I saw I was each time thinking it could be either a human or a computer...you know pretending to be a human...because obviously that’s the of course the subject of the experiment!”

[was it fun/booring...?]

“It was FRUSTRATING! I really wanted to ask him some questions related to other fields to try to (trick) him”

[what would you have liked to ask?]

“Oh, just...would have gone on to more abstract subjects...”

“I could have asked...well obviously talked about human feelings...see if he, the respondent, has some kind of sensibility...humour...I was trying to ask a few questions about humour because that was still more or less within the subject...I was trying to look for some kind of repetition in the answers, if he...and one of the respondents was basically saying the same thing like one or two or three times – the simple answer one – ...but, the other think is that I was thinking the other human would probably be trying to be a computer...I did not do it (laughs) but if he was pretending to make mistakes...”

Group 3

- Anna

“I thought it was very interesting. It was a bit difficult to talk about April Fools day so I didn’t know what to say and do...didn’t think April Fools day is that a special day...”

[mentions initially that she found it weird that they were using MSN for the experiment – but that as soon as she started and was trying to deduce who was who she did not question the set-up at all.]

“I think so, I think XX was the computer! I was talking about a bit odd prank to hide the fridge from my friend and I got a weird answer from XX...And also the spelling was wrong”

- Colin

[Appears very firm and calm. Tells me how his suspicion built up during the experiment]

“Started off by greeting and see if I got an answer. Tried to ask challenging questions. Questions that would challenge the respondents: both in language understanding and in logic”.

[Describes how he asked both for example if April Fools day is the day before the day after April Fools day. Neither answered this. The respondents just idled until he came up with a new question]

“First when I saw a spelling mistake I thought YES, this is a person...but then I thought about the TT and that you can tell a computer to make spelling errors so it would try and disguise itself – I got one spelling mistake and I doubted that a person would type in this (the same) spelling mistake all these times”

“The delays both made me think that it was a person typing an answer or that you had a programme with a delay function”

“Don’t know if I was supposed to do this: did not try to have a conversation but to find out!”

[a particular moment]

“Another suggestion was to move the fridge and hide the fridge, from his girlfriend, so that she would think the fridge was stolen...I told the other that that would be bad because the groceries would get damaged. So...

It was weird. It said that damaging the fridge would be ok for a good laugh. So it kind of didn’t think of the groceries – didn’t think of it deeper than just the fridge itself.”

[convinced he was talking to two computers]

- Ben

“Strange at first!...I didn’t know what to expect...I just started talking...didn’t try to detect just went open-minded”

“I wasn’t sure but I think I figured it out because I know my friends sense of humour! [laughs] Didn’t take long because I have a special sense of humour and so do my friends”

Group 4

- Carl

[Describes how he went about the experiment in detail] [when sitting in on this experiment I write down a note about that this participant kept saying “weird”, “weird” repeatedly.]

“One thing that I noticed right away is that one of them always start with a capital letter and ends with a period – which most people don’t do when they are chatting on the instant messenger thing “ [but then remembers that he has an older aunt and she always does this, so realises that it could still be a person]

“A question came up Where are you from...” [and from this conversation the participant realises who he is talking to regarding one of the respondents – whereupon he terminates the experiment]

- John

Describes the practical set-up and what happened carefully. Says in a calm friendly way that it was frustrating to have to wait for the responds so long.

“I was trying to work out which was the computer and which was the human, but I did not get many answers – only six responses!”

[Nevertheless however has “some ideas” about which was a computer and which a human but does not describe what these ideas are based around]

- David

“I just wait for long time and there is no response and I thought, yeah: maybe it is...I am talking to some machine!” [So you thought that it was one machine or two machines, or?] “BOTH!” [You thought both respondents were machines?] “Two machines! Yes! First...I thought I was talking to robots...because I have experience of talking to robots before several times...so, I thought: Yeah! This is it!” [snaps his fingers very definitely, confidently]

[Was also looking for odd answers. Read spelling mistakes as proof of human activity rather than computer activity.]

[Due to the fact that the conversation turns to where people are from we then quickly changes his mind and realises that he is talking to two humans!]

10.2 Part three: Thoughts in Retrospect About the Experiment.

Group 1

- Eric:

“I could have figured that!...can I guess now? Well, neither was an AI”,
 “Yeah, I felt the answers were...it felt like both were trying to be an AI not trying to chat!”,
 “...it was quite weird of an AI to say quotes Star Trek. I asked it how it knew and it answered April Fools were dummies. I didn’t think that AI would go that deep.”

- Emma:

“(Giggles. Silence) It’s a joke?” [I respond no, it is not a joke]
 “OH YEAH, I KNOW IT (laughs loud) I think I was talking to both guys! WAS I?! (laughs) I DON’T BELIEVE IT (laughs)” [before I have confirmed that this was indeed the case].
 “ Yeah, OK [after I have confirmed] ...it’s weird!!! (laughs) Actually I COULD believe that one was a machine and one was a person”

- Thomas

“(laughs) I am thinking you are going to tell me something funny now!”
 “Yeah, could be both computers and both humans! Could be.”

Group 2

- Clara

“Were they both computers!” [said as a suggestion]
 [when I ask the “if I tell you” – question she quickly says that she thinks it is more likely both are humans because of the non-formal language from one of them]

- Hugh

Laughs!!! Giggles. [Was very firm about which was a computer and which was a human. When asked the “If I tell you...” he says that he thinks he could have guessed and is certain it could NOT have been two computers – but says it could have been two humans]
 “when you ask a lot of questions and they come back in the right order...I am not sure a machine could have done that”

- Jim

“Obviously someone has been fooled! Well, I just have to figure out who! Well, I suppose it could have been four humans talking to each other! (laughs). And no computer at all!”
 [Could it have been two computers?]
 “Yeah, obviously they _could_ ...because I have no idea how far, how evolved these computers could be”
 [Could it have been two humans?]
 “...it would be possible...but again, this you come back to...this simple answer respondent...uhum...the attitude was just a bit strange...and I don’t think...I think that humans participating in this experience do get somehow excited, because otherwise they would not be here!...and that respondent looked very boored and not wanting to be there...and I thought...that did not make any sense...so...I don’t think he could be a human!”

[I tell him – he starts to laugh but looks very serious. I ask him if he is disappointed...]

“NO – I am relieved”

“I am relieved to see that computers are not that smart!”

“it was disturbing to see that one of these seemingly intelligent entities was a computer. To actually talk to one of these things is..._disturbing_!”

“It is very fascinating with robots, it is also disturbing...”

Group 3

- Anna

“A bit sneaky! Well...nah!”

[when I ask the “if I tell you” she hesitates and thinks and answers after a while:]

“...Aah, I don't know...maybe like 50-50...!”

- Colin

[Interview one. I reveal the true nature of the experiment]

“Ok. Ok.”

[Interview two. Colin reveals now that he did not actually believe me when I told him that he had been talking to two humans – he was simply so convinced! States that he could have gone on evaluating for an hour and he would have been sure to have come up with the same deduction – he had checked so carefully and in his belief come to such a solid conclusion]

[describes what went through his head when I had told him the true set-up]

“I was...surprised and...kind of started re-evaluating all the things that I , think about the things I, that led to the conclusion that both were computers...but for some reason in my head I was still certain that those WERE computers. I almost did not believe that those were not computers. All my reasoning told me that they were computers.”

“One of the things I thought about was, which is my own belief is that the TT is not a good thing to test AI. This - I was kind of happy it didn't work out! “ [laughs quietly/happily]

- Ben

“No way two computers! I know my friend's sense of humour! ...Two people...yeah, possibly...”

Group 4

- Carl

Sounds doubtful when I probe regarding the name of the experiment and if this makes him think in any particular ways. “no, I have no idea...which is probably why it is called april fools!”

- John

“Would they have been both computers?!” [No] “Both humans?!”

- David

[Discusses with detail how easy it is to hear that you are talking to a computer rather than somebody real. This is his experience from having done this via the internet. Relates to how this experience was on his mind all the time during the experiment]