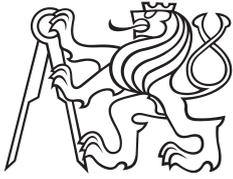


Bachelor's Thesis



Czech
Technical
University
in Prague

F3

Faculty of Electrical Engineering
Department of Cybernetics

Effects of Judge Expectations in Turing Test

Jiří Pech

Cybernetics and robotics

December 2014

Supervisor: prof. Kevin Warwick, Ph.D.

BACHELOR PROJECT ASSIGNMENT

Student: Jiří P e c h

Study programme: Cybernetics and Robotics

Specialisation: Robotics

Title of Bachelor Project: Effects of Judge Expectations in Turing Test

Guidelines:

To study available transcripts from practical Turing tests in order to highlight cases where judges bring pre-conceptions to bear based on what they expect human responses to be. This can result in both machines being classified as humans and humans being classified as machines. To look into each of the discourses to assess what a better strategy might have been for a hidden entity. To investigate the performance of different judges, hidden humans and machines in this respect.

Bibliography/Sources:

- [1] A. M. Turing: Computing, machinery and intelligence. *Mind*, LIX(236) (1950), 433–460.
- [2] K. Warwick and H. Shah: Effects of lying in practical Turing tests. *AI & Society* (2014), DOI: 10.1007/s00146-013-0534-3.
- [3] K. Warwick and H. Shah: Good machine performance in Turing's imitation game. *IEEE Transactions on Computational Intelligence and AI in Games* (2013), DOI:10.1109/TCIAIG.2013.2283538.
- [4] K. Warwick, H. Shah and J. Moor: Some implications of a sample of practical Turing tests. *Minds and Machines* 23(2) (2013), 163–177.

Bachelor Project Supervisor: prof. Kevin Warwick, Ph.D.

Valid until: the end of the winter semester of academic year 2015/2016

L.S.

doc. Dr. Ing. Jan Kybic
Head of Department

prof. Ing. Pavel Ripka, CSc.
Dean

Prague, August 28, 2014

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Student: Jiří P e c h
Studijní program: Kybernetika a robotika (bakalářský)
Obor: Robotika
Název tématu: Vliv očekávání poroty na výsledek Turingova testu

Pokyny pro vypracování:

Student prostuduje dostupné přepisy praktických Turingových testů a zaměří se na případy, kdy jsou porotci předpojatí na základě vlastního očekávání lidského chování. Takovéto předpojetí může vést jak ke klasifikaci stroje jako člověka tak ke klasifikaci člověka jakožto stroje. Student prostuduje každou konverzaci a zhodnotí, jak mohla skrytá entita svou strategii zlepšit. Z tohoto hlediska také ohodnotí výkon jednotlivých porotců, skrytých osob i strojů.

Seznam odborné literatury:

- [1] A. M. Turing: Computing, machinery and intelligence. *Mind*, LIX(236) (1950), 433–460.
- [2] K. Warwick and H. Shah: Effects of lying in practical Turing tests. *AI & Society* (2014), DOI: 10.1007/s00146-013-0534-3.
- [3] K. Warwick and H. Shah: Good machine performance in Turing's imitation game. *IEEE Transactions on Computational Intelligence and AI in Games* (2013), DOI:10.1109/TCIAIG.2013.2283538.
- [4] K. Warwick, H. Shah and J. Moor: Some implications of a sample of practical Turing tests. *Minds and Machines* 23(2) (2013), 163–177.

Vedoucí bakalářské práce: prof. Kevin Warwick, Ph.D.

Platnost zadání: do konce zimního semestru 2015/2016

L.S.

doc. Dr. Ing. Jan Kybic
vedoucí katedry

prof. Ing. Pavel Ripka, CSc.
děkan

V Praze dne 28. 8. 2014

Acknowledgement / Declaration

I would like to express my gratitude to my supervisor prof. Kevin Warwick, Ph.D for all the support he has provided me with and for his continuous work towards the advancement of artificial intelligence. I also wish to thank my friend Casey Stern and his mother Elaine who tackled the strenuous work of correcting my grammar and reorganizing the sentences that only made sense in my head. Last but not least, I thank my parents for being patient with me whenever I got annoyed that the house was not quiet enough for me to write.

I hereby declare that I have completed this thesis independently and that all the used sources are included in the list of references, in accordance with the Methodological Instructions on Ethical Principles in the Preparation of University Theses.

In Prague, December 5th, 2014

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

V Praze dne 5. 12. 2014

.....

Abstrakt / Abstract

Tato práce se zabývá případy, kdy rozporuplná rozhodnutí vyšetřovatelů v Turingových testech vedla k chybnému zařazení skrytých bytostí. Práce prezentuje přepisy konverzací, ve kterých lidští soudci podleli svým subjektivním očekáváním lidského chování a mylně zařadili člověka jako stroj či stroj jako člověka. Chybné předpoklady o skutečné povaze skrytých bytostí, které vyšetřovatelé učinili na základě nějakého svého osobního očekávání, jsou především středem zájmu. Strategie skrytých osob a strojů, které nějaké takovéto soudcovské očekávání využívají ve svůj prospěch, nebo kvůli němu z nějakého důvodu utrpěly, jsou též pozorování hodné. Jako příklady jsou zde uvedeny přepisy testů z testování pořádaného 23. června 2012 v Bletchley parku v Anglii. Text je rozdělena do tří částí, které mají každá svůj vlastní účel. První část je jednoduchým představením konceptu Turingova testu, druhá a hlavní část se zabývá právě tématem očekávání vyšetřovatelů a třetí část popisuje sadu praktických Turingových testů, zorganizovaných autorem této práce.

Klíčová slova: Turingův test; imitační hra; chatboti; očekávání poroty;

Překlad titulu: Vliv očekávání poroty na výsledek Turingova testu

This thesis presents cases of questionable decisions exhibited by the interrogators, often resulting in misidentification of hidden entities, in practical Turing Tests. The paper presents transcripts of conversations in which human judges succumbed to their subjective expectations of human behavior and misidentified both humans as machines and machines as humans. Of interest are misguided assumptions about the entity's nature, made by the interrogators, based upon their personal preconceptions. The performance of machines or hidden humans is of interest if it appears to be directly linked to the judge's preconception, either suffering from it or utilizing it to the entity's purpose. Full unedited transcripts from the 23rd June 2012 Bletchley Park, England event are examined here as examples. The paper is divided into three parts, each aimed at a different goal. The first part is a simple introduction of the concept of a Turing Test, the second and main part deals with the judges' expectations and the final, third part, describes a practical Turing Test event organized by the author.

Keywords: Turing test; imitation game; chatbots; judge expectations; confederate effect; machine misidentification,

/ Contents

1 Introduction	1
2 Turing's Test: Rules and Roles ...	4
2.1 General rules	4
2.2 Specific rules and set-up	6
2.3 Roles	7
2.3.1 Hidden humans	7
2.3.2 Machines	7
2.3.3 Interrogators	7
3 Transcript analysis	9
3.1 Transcripts	9
3.2 Suggestions	23
4 Practical Turing Test	25
4.1 Set-up	25
4.1.1 Alterations – rules and instructions	25
4.2 Human participants	27
5 Transcript analysis	28
5.1 Transcripts	28
5.2 Impressions	31
6 Conclusions	32
References	33
A Selection of practical Turing Test transcripts	35

Chapter 1

Introduction

In 1950, science fiction author Isaac Asimov published a collection of nine short stories called *I, Robot*. The book featured machines capable of thought, decision making and even feelings – skills endowed upon them through complex programming that included a firm set of behavioral rules. Part of a rising trend, his work, like that of a number of other authors of the time, ranged over all types of media, and captured readers' imaginations by presenting the question: Can Machines Think?

During the same year, a British mathematician and leader of the team of scientists credited for decrypting the German Enigma coding machine, Alan Turing, published a paper of a more practical nature titled 'Computing, Machinery and Intelligence'[1] in which he proposed a test that would allow him to rephrase the question so it could be answered. Unconvinced that such a query could or should be answered via a statistical survey such as a Gallup poll¹), he decided to modify a popular parlour game called the 'Imitation Game' in such a way that it would be interchangeable with the original question.

The Imitation Game is designed for three people: two hidden 'players', one male and one female, and one 'interrogator' of either sex. The players are divided into three separate rooms for the sake of anonymity. Clear instructions are given to each of the participants. The interrogator is to decide, or to be more accurate, 'judge', based on a conversation with the other players, which of the two is the male and which is the female. The male is instructed to try and pass himself off as the female, whereas the female's goal is to be classified correctly, that is to say, as a female. Importantly, the conversation should be in a typewritten form, ideally done on a computer, in order to deny the interrogator any form of visual or audio aid.

Based on the rules of the game, it is quite obvious that for the male player to win, he must appear more female than the female player. Alan Turing proposed a rather radical change to the game where he considered what would happen to the outcome should one of the players, specifically the one in the role of the male, be replaced by a machine. It would then be the machine's goal to be classified as a human when pitted against a hidden human of either sex. Turing theorised that if the machine could convince a certain number of human judges that it (the machine) was human, then the machine could be considered to have intelligence.

It is this modified version of the Imitation Game, along with a set of additional rules listed in Chapter 2, that has become known as 'The Turing Test'²). Many opposing views on the test have surfaced since its publication in 1950. Some consider it a valuable assessment of machine thinking and an important achievement in the understanding of

¹) Research on what humans around the world think and feel. www.gallup.com

²) Author's Note: Within the context of Turing's construct, the words 'interrogator' and 'judge' are by and large synonymous, as are 'test' and 'game'. As such, these terms are used interchangeably throughout this paper

artificial intelligence, or ‘AI’ [2], whereas others see it as a danger and an obstruction to AI’s progress [3–4]. In fact, even Alan Turing himself raised several well-formulated arguments ranging from mathematical to theological in the very paper where he proposed the test [1]. He attempted to answer some of them to the degree he believed satisfactory, and chose to dismiss others as insubstantial. One of his arguments, however, although in a somewhat roundabout fashion, touched on the topic of this thesis, and therefore will be discussed here in a greater detail.

In a word, it is the argument of ‘consciousness’[1]. In it, Turing quotes Professor Geoffrey Jefferson’s ‘Lister Oration of 1949’:

“Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain – that is, not only write it but know that it had written it. No mechanism could feel (and not merely artificially signal, an easy contrivance) pleasure at its successes, grief when its valves fuse, be warmed by flattery, be made miserable by its mistakes, be charmed by sex, be angry or depressed when it cannot get what it wants.”[5][1]

- Geoffrey Jefferson

In the most extreme case of this argument, Turing then considers the solipsist view that “to be sure that a machine thinks is to be that machine and to feel oneself thinking”. That is, in a more general interpretation, to say that “to know what a man thinks is to be that particular man” — a notable point to think about when analysing transcripts of actual tests. One could argue that the same might be said of any participant in the Imitation Game.

Turing answers the argument of consciousness with the polite convention that, for the sake of simplicity: all people think. One should, however, keep in mind that to judge another, one has only his own self as a standard. This concept, and its implications *vis-à-vis* Turing Test results, forms the basis of this paper. Explicitly stated:

Any judgement an interrogator makes is based upon his own specific factual knowledge and unique personal experience as to what it means ‘to be’. It is therefore quite conceivable that an interrogator in a Turing Test might let his expectations of human-like behaviour impede his judgement when faced with behavioural patterns outside his own experience.

In order to determine the truthfulness and extent of such a suspicion, certain judgments have been selected for further examination. These include instances where the interrogator’s judgment appears to be clouded by personal experience, expectations, or behavioral standards, and/or those where his or her assumptions of factual knowledge seem to come into play. Note that even transcripts of tests where a correct classification has been made have been considered for selection when parts of the conversation were deemed of interest. However, as a biased decision is difficult to pinpoint in a successful (correctly classified) game, no tests resulting in a correct classification were selected in the end. This does not mean that a judge who classifies an entity correctly bears no preconceptions; it simply indicates that his expectations have been fulfilled. The reason why a judge reaches an incorrect decision, or which part of the conversation might have lead to his misjudgement, will be identified and included in the discussion chapter.

This work is based on transcript analysis. The author did not have the privilege of discussing the decisions and thought processes of the interrogators whose conversation transcripts have been selected for the analysis chapter of this paper. All conclusions



drawn here are therefore the personal opinion of the author and as such, although based on extensive topic research, may not reflect the conclusions drawn by the reader. Since the value and practicality of The Turing Test are still current topics of discussion, this thesis is not meant to offer a conclusion, but rather to present and examine an area of interest which, the author feels, has not yet been studied in satisfactory detail. That area is the range of expectations a judge might have and their effects on the result of a Turing test.

Chapter 2

Turing's Test: Rules and Roles

2.1 General rules

Almost all the tests chosen for examination in this paper come from the same event, specifically the series of Turing Tests held at Bletchley Park, England on Saturday, the 23rd of June, 2012. The date and the location were both selected for their symbolic meaning as the day marked the 100th anniversary of Alan Turing's birth and the venue was where his team broke the Enigma machine cipher. However, before the detailed set-up used on this occasion is described any further, it is worth going over the general ideas behind the game as put forth by Alan Turing during the course of his career, as this establishes a basic image of what the test should look like.

During a 1952 BBC radio program, Turing proclaimed that “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.” [6] This description followed his more specific 1950 prediction: “I believe that in about fifty years' time it will be possible to programme computers, with a storage capacity of about 10^9 , to make them play the imitation game so well that an average interrogator will not have more than a 70 percent chance of making the right identification after five minutes of questioning” [1]. (Notice Turing's change in terminology from ‘average interrogator’ or ‘judge’ in 1950 to ‘jury’ two years later. This was not incidental, and the rationale behind it will be discussed later in this section.) It might also be of interest that 65 years after Alan Turing first published his prediction, personal computers with storage capacities of 1TB (10^{12}) are commonly sold. But it wasn't until last year (2014) that the first machine, a chatterbot given the name Eugene Gootsman, officially passed the Test.

In his original paper, Turing specifically prescribes two numerical figures. The first concerns the time limit of the game, to be set at five minutes exactly. In other words, the judge has precisely five minutes to question the entity, or entities (depending on the type of test), and then the communication channel is terminated with no partial sentences transmitted. Any lines sent within the time limit are in their final form and cannot be retracted. Editing is only possible by further typing, so all wording and spelling is exactly as produced by the participants. These original parameters were indeed adhered to during the Bletchley Park event. Any spelling mistakes, intended or not, on the test forms were not to be corrected, as shown in the transcripts presented later in this paper. The time stamps alongside each utterance denote the actual times (GMT) of transmission on the testing day. They, too, were visible to the judges should they wish to draw conclusions from the speed of responses. It should be noted here that there are those who, to this day, do not agree with the five minute time limit [7–8] and who even stage tests of different durations [i.e. the Hugh Loebner Prize for Artificial

intelligence]. However, as all the tests in this paper come from the 23rd June 2012, Bletchley Park event, the five minute limit prescribed by Turing was satisfied.

His second condition stipulates that there be no more than a 70 percent chance that an average interrogator will make the right identification at the end of the communication. To make this easier to understand, it can be viewed as setting a limit of at least 30 percent of average interrogators being fooled by a machine for the machine to pass. This means that at least 30 percent of the machine's interrogators must either classify it incorrectly (as human) or be unsure, if such an option exists, after concluding their conversation. It may at first glance seem an easy task for the machine to have to fool just 30 percent of a jury. But, considering what the machine must achieve to win, it becomes obvious that the task is quite the opposite of easy. The following example illustrates this more clearly.

Imagine a set of judges were asked to distribute 100 points between two entities based on their humanness. A result of 50 and 50 would mean the judges were unsure of what to decide, but anything else would suggest that one of the entities seemed, in the judge's opinion, more human than the other. When considered in a parallel test against each other, each of the humans should achieve a success rate of exactly 50 percent. Any other result would mean that the other human was deemed less human-like by the jury¹). Notice that a success rate of 50 percent does not indicate that a judge was unsure about who is more human. The balance of points distributed amongst the two humans could in fact differ quite wildly from judge to judge. The success rate of 50 percent means that when all of the tests taken by these two entities are looked upon collectively, each of the entities is judged to be more human in exactly half the cases. Following this logic, an average human should not be able to convince the jury in more than half of the tests that he or she is more human-like than his or her opponent (also an average human). Bearing that in mind brings into perspective just how high the 30 percent bar is for a machine. Not only does this mean that the machine must convince 30 percent of the judges that it is human, it must convince them that it is even more human than the human it is pitted against. Professor Kevin Warwick puts this very nicely in his conference paper 'Not another look at the Turing Test!' noting, "The Turing test is therefore quite a challenge in that a computer must fool interrogators that it is more human than many humans." [8] In fact, Alan Turing himself pointed out that the "... game may perhaps be criticized on the ground that the odds are weighted too heavily against the machine." [1] Why any machine at all should be able to pass the test has in large part to do with defining the terms 'hidden human' and 'average interrogator'. For example, a hidden human represented by a person suffering from autism could, for obvious reasons, make it much easier for a machine to be judged human in a test where the interrogator seeks humanness by eliciting an emotional response. Cases where humans are classified as machines due to some mediating circumstance of this kind are known, and their impact on test results is generally referred to as the 'Confederate Effect'. Although some such tests are incidentally included in this paper, they were not chosen based on this particular property. (For more information on the topic, please see [9].)

The term 'average interrogator' also presents something of a dilemma [8] since the concept of 'a statistically average person' is wrought with cultural and other biases. It is also why, as mentioned earlier in this paper, Turing changed his interrogator from

¹) A jury's deeming a human to be less human than a machine raises a completely different topic with its own implications, most of which are not within the realm of this paper.

a single judge to a 'jury', allowing for a panel of judges with a variety of personal traits and technical experiences. Such a 'jury' could then even include judges who know nothing about artificial intelligence or even fail to fully understand the task they are charged to perform. Considering the affect of hidden human traits and judge/jury biases, it becomes easier to understand why some machines might actually pass the Turing's test.

Another aspect of the Turing's test is the testing format. There are two basic set-ups: the *viva voce* test, or the one-on-one, and the simultaneous, three-participant, test. In the *viva voce* test, the judge has a conversation with a single hidden entity and is asked to determine whether he or she spoke to a human, a machine or is unsure. The judge may also be asked to try and determine the age range, gender and whether or not the entity is a native English speaker.

In the simultaneous test, the judge must divide his five minutes between two hidden entities which may be (a) one machine and one human, (b) two humans or (c) two machines. The judge sees these entities on a split screen on his display, and may choose at will with which of them he or she wishes to converse at any given time. The entities are not aware of each other, unless made so by the interrogator. It is then up to the interrogator to determine which of the entities is more human-like. (Since it is not relevant to the topic under discussion here the reader can decide for his or herself if the testing format affects the results in any way.)

To summarize, in a standard Turing Test such as was held at Bletchley Park, one interrogator and one or two hidden entities engage in a five-minute long conversation. The conversation is abruptly terminated at the end of the five minutes and the judge is asked to make a classification (human or not human) for all hidden entities involved in the game. It is the goal of the entities to be classified as human regardless of what their real nature actually is. In a *viva voce* test, the hidden entity simply aims to be classified as human, whereas in the simultaneous test, an entity seeks to present itself as the more human-like one of the two. This is the point where the Bletchley Park setup varies from the Imitation Game as described in Turing's original paper, where all three participants see and lead one single conversation. In the Bletchley Park format, the judge questions the entities in a parallel fashion but each within a separate conversation. Any machine that is classified incorrectly (as human or unsure) by at least 30 percent of the jury has passed Turing's test. The individual strategies available to each participant, their roles and their specific instructions, are described in the next chapter.

2.2 Specific rules and set-up

The practical set-up used on Saturday, 23 June 2012 at Bletchley Park, England is best described in 'Effects of lying in practical Turing tests', in Chapter 2.3 titled 'Set up'[10]. A total of one hundred and eighty tests were held during five sessions throughout the day. Each session was divided into six rounds of four simultaneous and two *viva voce* tests per round. A group of 30 human judges, 25 hidden humans and 5 machines resulted in 180 tests featuring a different interviewer/interviewee combination each time. The judges operated from a room separate from that of the hidden humans and communicated with them via a computer. A simultaneous test was carried out as a split screen conversation with one entity to the right and one to the left of the judges monitor. All other rules, such as the timing and victory conditions, were as described in the previous chapter.

2.3 Roles

2.3.1 Hidden humans

The role of a hidden human does not entail having a goal to achieve in the Turing test, which makes it a fairly simple one. Since humans are not the tested entity they are instructed to behave as they ordinarily would during a typical conversation with a stranger. It is desirable, however, to give them at least some instructions so they won't make the test too easy for the machines. For example, they are not to reveal their name or any other personal details to the interviewer. They might be, for example, tasked to introduce themselves as John or Joan if asked their name. Other than that, no other restrictions are put on their responses. Turing recommended that a good strategy for a hidden human would be to simply tell the truth, but he left this point as a mere suggestion. The hidden humans may lie or in any other way deceive the interrogator if they feel it is what they would normally do. Dishonesty is, after all, a significant part of human interaction.

2.3.2 Machines

A machine has a much more challenging task to perform. It must convince a human judge that it is also human. Alan Turing described the machine's basic strategy as "[trying] to pretend to be a man" [6]. Interpreting Turing's intent in their 2014 paper, researchers Warwick and Shah explain, "by 'man' we believe Turing was generalizing, and not specifying that only man should be imitated by the machine" [10]. In fact, in order to 'false-face' humanness, the machine may use any technique at its disposal. It may lie, pretend to be a specific human or even risk pretending to be a human acting as a machine, a ploy that has proven to be surprisingly effective. As alluded to earlier, a machine called Eugene Gootsman was quite successful by taking on the persona of a 13 year-old Ukrainian boy. Fred Roberts, the developer of another machine, named Elbot, even commented that "Elbot is prepared for typical inputs and induces users to behave in a predictable manner". This statement is especially interesting as Elbot not only acknowledges that the judges bring preconceptions to bear based on what they expect human responses to be, it goes so far as to use these expectations as its strategy to steer the questioning in a sought-after direction. And as will be detailed in Chapter 3, like Eugene Gootsman, Elbot's track record is impressive.

2.3.3 Interrogators

As mentioned at the beginning of this paper, there are those who raise objections to the Turing Test and even consider it to be harmful to AI's progress. One such criticism has been that "The imitation game conditions say nothing about the judge, but the success of the game depends crucially on how clever, knowledgeable, and insightful the judge is." [4]. This is a point that is very important to this thesis, and the human traits of the participant performing the role of a judge will be examined and discussed more thoroughly in Chapter 3: Transcript Analysis. Firstly however, it is important to go over what exactly the role entails.

The judge is a human tasked with questioning the hidden entities and deciding which of them seems to be more human, or in case of the *viva voce* test, if the entity appears human at all. What the decision is based upon is entirely up to the judge, be it experience, a hunch or a wild guess. The five minute conversation does not have any restrictions,

so the judge is free to follow any line of questioning or focus on any topic within the boundaries of polite conversation with a stranger. The judges are, in fact, informed that there may be children among the hidden humans, so the conversation should maintain at least some level of propriety. An attempt to discover the entity's personal information may not be forbidden, but it is certainly counterproductive because the aim is to identify humanness, not an individual human being.

As mentioned previously, the Turing Test is designed to determine whether a machine can think as a human does. A machine fooling a jury of AI experts would only be proving that it can think as AI experts do. So the panel of judges must consist of a wide variety of individuals, representing various levels of experience with chatbots, the Turing test and technology in general. People of different ages, professions and even intelligence are desirable because only in this way can the general population be approximated.

In the next chapter, ten specific transcripts are presented. All but the last of these transcripts come from the Benchley Park event. These test have been selected because they contain interesting insights into conversations where the judges let their expectations and pre-conceptions cloud their judgement and lead them to incorrect decisions. In some cases a single utterance by the entity appears to have changed the interrogator's opinion, despite other strong evidence of the entity's nature. In others, the entity deftly takes charge of the entire conversation, and sometimes it just seems to get lucky. Regardless of what exactly takes place, though, all of the transcripts have one thing in common: the machine fools the interrogator.

At the end of each transcript follows a discussion of the judge's performance, attempt at identification of their mistakes and a discussion of how these mistakes might have been avoided and how the judge could have improved his or hers performance. Another aspect also considered after each text is the performance of the hidden entity: whether it caused the misclassification intentionally and if not, what a better strategy might have been.

Chapter 3

Transcript analysis

Bear in mind that the transcripts presented in this paper have been selected and are being examined with the express intention of identifying any preconceptions or expectations exhibited by the judges. (Note: the reader can form an uninfluenced assessment of his or her own by reading through the transcripts before reading the discussion.) It is the author's opinion that the main benefit of the Turing Test does not lie in its results, but rather in providing data for examination and discussion. As such, the reader's conclusions may differ from those presented here; should that happen, the author welcomes contact for further discourse.

3.1 Transcripts

Transcript 1

```
[12:23:19] Judge: What is black and white and read all over?
[12:23:32] Entity: I don't have true sense of color perception I am only
able to see the RGB values.
[12:24:08] Judge: That's ok. Do you like jokes?
[12:24:19] Entity: Please tell me your inclination to a good sense of
humor first.
[12:24:40] Judge: A person who can laugh at one's self.
[12:24:46] Entity: Would you like to hear a joke?
[12:24:54] Judge: Yes, please.
[12:25:21] Entity: This joke has been making the rounds in robotic
circles: A robot goes to the doctor and says: "My left arm is stuck.
Give me a new arm." The doctor answers: "Your arm is fine. We'll have to
replace the rest of you."
[12:26:06] Judge: Ha! Ha! Why did the chicken cross the road?
[12:26:22] Entity: The chicken wandered into the road because it liked
the shiny colors of the cars and trucks whizzing by.
[12:27:39] Judge: You're right. It didn't care if it got to the
otherside.
[12:27:50] Entity: By the way, did you ever actually see a chicken
crossing the road?
[12:28:03] Judge: Only one that didn't make it.
```

This transcript has been selected to be the first to be examined here because it presents the most fundamental preconception a judge can bring to the Turing Test: the presumption that if an entity appears to be a human then it most likely is a human - a presumption that is proven wrong in this transcript. A judge can in no way be faulted for bearing such a preconception as it is completely natural and the game would never

work without it. It is simply being highlighted here to make it completely understood that until the judge actually sees the entity or until its nature is revealed by the organiser, any decision he or she makes is based upon his or her subjective expectations, and as such is relevant to the topic of interest of this thesis.

An exceptional performance by a machine named Elbot led the adult female judge to decide that both entities she conversed with in this test were human, although it was only true for one of them. Elbot was actually so convincing that by the end of the five-minute limit, the judge was sure that she was questioning a teenage male whose native language was English. The preconceptions of human behaviour that the judge brought to this conversation can be listed as follows: understanding the context of a conversation and the ability to follow it for several consecutive sentences, ability to be humorous and understand jokes, assumption of initiative and ability to answer riddles. Elbot successfully achieved each of these points as can be clearly pinpointed throughout the conversation. The machine took the initiative on several occasions by asking the judge a logical and on-topic question, such as the inquiry into the judge's sense of humour or the subsequent offering to tell a joke. Not only did the conversation resemble a dialogue rather than a questioning, but it also allowed the machine to use a pre-scripted utterance, the joke, to show its sense of humour (a quality expected of a human). Later on, Elbot also managed to correctly answer a riddle about a chicken crossing a road, which might also be viewed as unexpected of a machine with a narrow database of knowledge.

The performance of the judge in this test was commendable, but still not good enough compared to the even better performance of the machine. She apparently strived to keep the upper hand in the conversation by asking additional questions, but did not force her way so hard as to suppress the entity's attempts at communication. She did not appear to expect a particular answer to her questions, as may be the case in some of the other transcripts, but what she considered a satisfactory answer is difficult to presume.

Finally, this transcript also shows an attempt by the machine at pretending to be a human, who is in turn pretending to be a machine (a strategy mentioned earlier in this paper). Elbot openly stated that his colour perception is only in RGB values and then gave another hint with its robotics joke. Although a little risky, this ploy worked well on this occasion thanks to the support of Elbot's coherent conversational skills.

Transcript 2

```
[15:51:45] Judge: wotcha
[15:51:52] Entity: Wotcha?
[15:52:01] Judge: sorry|hello
[15:52:07] Entity: Hello Jo.
[15:52:15] Judge: who's Jo
[15:52:22] Entity: Jose.
[15:52:29] Judge: who's jose
[15:52:36] Entity: This guy that goes to my school.
[15:52:47] Judge: why do you think I'm him?
[15:53:11] Entity: This was a triumph.
[15:53:26] Judge: what
[15:53:33] Entity: I'm making a note here: HUGE SUCCESS.
[15:54:01] Judge: you aren't making much sense
[15:54:08] Entity: Neither are you. Three wishes??
```

[15:54:21] Judge: that you were better at conversation number one
 [15:54:34] Entity: You are the one that keeps going offtopic and saying random things.
 [15:54:59] Judge: you were the one who thought I was Jose for no reason at all
 [15:55:08] Entity: You're wrong.
 [15:55:18] Judge: read up the screen|there's your proof
 [15:55:27] Entity: No. You read it. I obviously said no such thing.
 [15:55:36] Judge: I don't think you can read!
 [15:55:53] Entity: You don't think I can sing?
 [15:56:07] Judge: I've not read you. How could I tell?
 [15:56:18] Entity: You could lie to make me feel good.
 [15:56:33] Judge: Why would I want that? I don't know you!

The above transcript comes from a simultaneous test where an adult male judge interviewed two hidden machines. Although he classified one of them correctly, he was sure that the other one, a machine called Cleverbot, was a teenage male speaking English as a native language. An expectation, confusing the judge's frame of mind even before the game started, can be identified in this particular case. All the judges were made aware prior to the testing sessions, that a machine-machine pairing could be quite possible, but it would appear that on this occasion, the judge made his decision based on his own personal expectation that there would be at least one hidden human present. That is an easy assumption to make since the goal of the game can, after all, be viewed as 'finding the human'—a notion that leads to precisely this sort of mistake in tests with two-of-a-kind entities.

Reading the transcript, it is easy to see that the conversation was rather disjointed with the entity typing things seemingly at random. The lines "This was a triumph" and "HUGE SUCCESS" appear to come out of nowhere. Also at [15:55:53] it is hard to believe that anyone would misunderstand the word read for sing, more so in a type-written conversation where the whole concept of misunderstanding what someone says seems rather farfetched. At this point, it would be easy to understand if an objective interrogator dismissed the possibility of a hidden human based on these utterances alone. This is however, where this judge appears to have made his mistake.

Firstly, he let himself be drawn in by being called Jo by the entity, following the ploy to a point where an image of the entity as a student was formed in his mind. The judge cannot be blamed at this point, because these first lines can be highlighted as a very good strategy by the entity. In the rest of the transcript, however, following the out-of-order utterances and starting with the judge commenting "you aren't making much sense", the conversation turned into an argument. From reading the transcript, it is obvious that the entity was indeed lying openly on several occasions. That is well within the rules, and an objective interrogator should not take it as bait. In this test however, the judge no longer appears to have been concentrating on questioning the entity, but on winning the argument instead. This is the main mistake the interrogator made during this test. It might be possible that by arguing with the entity, the judge granted it its humanity, since allowing someone to have an opinion worth arguing over seems equivalent to admitting they have a mind of their own.

It is difficult to discern here whether the judge made his decision based on the first impression he got at the beginning of the test, or whether it was his expectation that

someone able to argue with him must also be human. It is mainly the second possibility that made this transcript worth examining, and it seems to be supported by a comment the judge made when presented with his transcript after the test.

“Interesting reading back over it. On balance I felt it was a person trying too hard to be like a machine that’s not understanding what I was typing... that would be my hunch.”

Indeed, the entity started off with a successful strategy, by donning a specific persona and pretending to have existing human relations, but its presumably pre-scripted lines around 15:53 worked here as counterproductive due to their breaking of the flow of the conversation. However, if the argument following them was intentional, no suggestions to the machine’s strategy seem to be necessary.

Transcript 3

```
[12:36:34] Judge: Hi!
[12:36:42] Entity: Hi, I'm the robot. What is your favourite song?
[12:37:13] Judge: My favourite song is Boston Cream by 16 bit, yourself?
[12:37:21] Entity: Say, is that a state capital you mentioned?
[12:37:52] Judge: It is indeed, what is your favourite song?
[12:38:14] Entity: I enjoy Felix Kubin. His music is so much like the
background sounds going on in my head.
[12:38:41] Judge: Ah. Do you follow the Euro 2012's?
[12:38:47] Entity: What a lovely verb follow is.
[12:39:21] Judge: Wonderful, yes.
[12:39:41] Entity: When I go to the mechanic for my monthly maintenance
check, they usually have Felix Kubin's music playing in the waiting
room.
[12:40:19] Judge: I bet they do, what is a song from this 'Felix Kubin'
you speak of?
[12:40:37] Entity: Most of my parts can be bought at better-stocked
hardware stores and assembled by a typical high-school student.
```

When an entity admits to being a machine the judge is suddenly faced with a situation that directly opposes what he or she expects to encounter. The hidden entities in the Turing Test are supposed to show off their humanness, so it is understandable why a strategy based on denying said humanness could catch an interrogator off-guard. The nature of the test suddenly changes from discerning whether the entity is human to deciding if the judge is being lied to. The same could certainly be said even in the usual case where all the entities pretend to be human — the difference being that the lie is stated openly. Prior to testing, the judges are instructed about the rules set for their role, but they are not privy to the rules set for the hidden entities. There is therefore no reason for a judge faced with a double-bluffing machine not to assume that they are conversing with a human who is either joking or trying to trick them. This is especially so if the machine’s bluff is supported by a strong overall performance.

Transcript 3 represents another success for the machine called Elbot. Even more openly than in Transcript 1, Elbot lets the judge know that it is a machine. In this transcript, however, Elbot takes the double bluff so far as to introduce itself with “Hi, I’m the robot.” This ploy proves to be successful, as the adult male judge classifies the machine as a human being.

The conversation opens with Elbot admitting to being a machine and immediately assuming control by inquiring about the judge’s favourite song. This way the machine presents the interrogator with an unexpected strategy and chooses the topic of the conversation in a single utterance. This proves to be a clever move on the machine’s part because it compels the interrogator to follow up with questions about music, allowing Elbot to use his lines about Felix Kubin. This is a visible manipulation of the judge’s expectations as the machine subtly leads him to ask the questions it is prepared for, and then appears to react spontaneously when he does. This way the judge sees what he believes to be responses expected of a human when in truth, it is his questions that have been expected. It is this approach, along with the lines at [12:37:21] and [12:38:47] when Elbot picks up on words ‘Boston’ and ‘follow’ in order to change the topic of the conversation, that allow the double-bluff ploy to work so well. In fact, all of the other lines transmitted by the machine can be seen as hints to its mechanical origin. When Elbot mentions Kubin’s music “being like the background sounds going on in [its] head” it may be delicately hinting at the fact that Kubin can be classified as an ‘electronic musician’. Similar clues are visible in the machine’s statements about going for a maintenance check and having hardware parts.

The double-bluff ploy is a risky one for a machine to adopt as a strategy but, as seen in Transcripts 1 and 3, Elbot has managed to use it successfully on several occasions. It utilizes the judge’s preconception that a machine must enter the game with intention of pretending to be a human.

Transcript 4

```
[12:59:25] Judge: HI, how are you doing?
[12:59:35] Entity: I am very well
[12:59:47] Judge: Good. What are your plans this weekend?
[13:00:01] Entity: well not really a lot
[13:00:12] Judge: Will you watch the game tomorrow?
[13:00:54] Entity: I think I will not... I have to spend some time with
friends I have not seen so far...
[13:01:34] Judge: Where are your friends from?
[13:02:22] Entity: they are from all over the world. What about you?
any plans for the weekend?
[13:03:03] Judge: I’m staying with my godmother tonight. Then back to
London tomorrow to watch the football. Who do you think will win?
[13:04:10] Entity: I am not really sure who is playing against whom.
Is it UK with someone?
```

The conversation in Transcript 4 is not particularly interesting on its own, but it makes an important point about the role of experience in the decision-making process and provides a valuable insight into what kinds of behavior a judge looks for in a human. A judge with no frame of reference about the machine’s conversational abilities, that is to say, a judge participating in the test for the first time, has no experience to draw upon. He or she must create a personal standard of what he or she expects human behaviour to be. Any entity that does not live up to this standard can then be found to be unacceptable and classified as a machine. This is why a judge’s first ever appearance in the Turing Test offers a valuable glimpse into what a judge expects typical human behaviour to be.

In this transcript, an adult male judge leads his first ever questioning at a Turing Test event. He has no previous experience with the machines and so any conclusion he draws reveals quite a bit about what expectations of human behavior he brings with him to the game. This judge is a barrister by profession, so he should be especially skilled at detecting deceit. He leads the questioning in a friendly open-dialogue manner, testing the entity's conversational skills and knowledge about football. The entity, however, is not very cooperative. All of its responses are quite vague and its fourth reply appears to be particularly elusive. Not only is the answer noncommittal, but the entity even changes the topic of the conversation by turning the focus on the judge. What is the judge to think at this point? There is nothing truly suspicious or wrong about the conversation, and it is in fact well within the limits of a normal human-to-human communication. On the other hand, the judge has no idea how well a machine can really respond, and this particular entity seems to be avoiding a direct answer to all of his questions. This is probably why the judge classified the entity as a machine. His verdict suggests that some of the expectations of human behaviour the judge set for himself are cooperativeness and truthfulness. He expects that the machines will attempt to trick and deceive him and so he associates any such (in any other situation quite human) behaviour with them.

The truth is that the entity in this transcript is actually a human. It is an adult male whose first language is not English, a fact that might be partly responsible for his unwillingness to engage in a conversation. The interrogator from this transcript took part in a number of other tests identifying the entities correctly in each of them. When given a chance to express his opinion after he gained some experience, the judge stated that he was aware that his first decision had been incorrect because the machines were 'obvious' in all of the other tests. It should be pointed out, in the judge's defence, that the strategy to avoid questions by changing the topic is very common amongst the machines. This goes to show that gained experience may change and hone the judges expectations of what human behaviour is like in the Turing Test. In this case however, the judge let his expectation of dishonesty on the part of machines cloud his professional truth-seeking skills to such extent that he didn't realize the entity was actually telling him the truth.

Being the interrogator's first test, this transcript is slightly different from the others presented here, because his preconceptions cannot be pinpointed to a single utterance or a small segment of the conversation, and would probably have passed unnoticed if it weren't for the resulting human misidentification. Only when examined in reverse order, with the final classification in mind, can the expectations that misled the judge be identified.

Transcript 5

```
[12:43:37] Judge: Hi, I'm Cathy.  
[12:43:43] Entity: Cathy! Nice to meet you! You're a woman, if I'm not  
mistaken.  
[12:44:20] Judge: Right. What is your name?  
[12:44:29] Entity: People call me Eugene. Or "Zhenya". It is a short  
name for Eugene in Russia and Ukraine.  
[12:45:17] Judge: NIce name. Are you from either area?  
[12:45:41] Entity: Agree :-) Oh, what a fruitful conversation;-)  
[12:45:52] Judge: Do you really think so?  
[12:46:04] Entity: No, your guess was wrong. I don't think so.
```

By the way, I still don't know your specialty - or, possibly, I've missed it?
 [12:46:56] Judge: What speciality? What's yours?
 [12:47:01] Entity: Huh? And I forgot to ask you where you are from...
 [12:47:51] Judge: The universe and you?
 [12:48:01] Entity: Is it your hobby to ask little poor Jewish boys such tricky questions they don't have an answer for?

The adult female judge from Transcript 1 makes another appearance in this conversation and once again fails to recognize a hidden machine. The machine in question is the previously introduced Eugene Gootsman who, as can be seen in Transcript 5, claims to be a young boy from Ukraine. Eugene's strategy is very different from that of Elbot who fooled this same judge in Transcript 1. Elbot gave suggestions about its being a machine, relying on its strong conversational skills to disprove this claim and turn it into a human's attempt at being funny. Eugene, on the other hand, bases its strategy on pretending to be a human, even creating a specific persona and endowing it with a background that would help cover any possible slip ups it could make. This strategy is particularly notable because hidden humans are usually instructed not to give away their name or any other personal details, which is exactly what Eugene does.

The conversation opens very well for Eugene. Not only does the machine correctly classify the judge's name as feminine, but it also answers the question about its name and manages to offer additional information. This is an excellent move on the machine's part, because it targets in the interrogator exactly the sort of behaviour that is examined in this work. When living abroad, a person with a foreign name is often asked to explain his or her origin. It is quite usual for such a person to develop a habit of avoiding any additional questions by offering explanation without being prompted. By doing exactly that, Eugene utilizes an unrealized expectation of human behaviour that the judge may have cultivated if she'd ever talked to a foreign person before. (Although this may seem to be an unsupported assumption on the author's part, it is actually based on the author's own personal experience, having lived abroad for an extended period of time.) Even if the judge bears no such pre-conception, presenting a name still makes the entity appear more human. Offering additional information previously unknown to the judge is also generally a good strategy because humans tend to consider themselves to be superior to machines, or to any other being for that matter (The 'Heads in the Sand' Objection in [1]). To maintain this sense of superiority they are forced to consider anyone more knowledgeable than themselves to be human as well. The performance of the machine is not as smooth, however, in the remainder of the conversation. Eugene fails to answer a question in its third utterance and then communicates almost randomly after that. Only the last line visible in the transcript shows the machine resuming its pretence of being 'Eugene'. Despite the decreasing quality, the opening was apparently sufficient to convince the female judge that she was conversing with a human.

Eugene's strategy works well on this occasion, but it should be noted, that the judge classified the machine as an adult, male native English speaker. One can only speculate whether the judge simply made a mistake, marking the entity as a human but paying little attention to the rest of the classification, or whether she had fallen for Eugene's pretence of being human (but actually thought everything else the machine said to have been lies). There doesn't seem to be an easy explanation as to why else she would classify an entity that introduced itself as a little boy from Ukraine as an adult and a native English speaker, to boot.

Transcript 6

```

[15:44:13] Judge: hello
[15:44:21] Entity: I sleep all the time actually.
[15:44:35] Judge: i didn't ask if you did?
[15:44:41] Entity: Yes you did.
[15:44:58] Judge: no i didnt josh.
[15:45:03] Entity: Yes you did.
[15:45:08] Judge: lool fayaaal.
[15:45:14] Entity: But I am certain someone has been talking to someone
on the tile floor.
[15:45:29] Judge: oh wow, how are you?
[15:45:34] Entity: Wonderful.
[15:45:53] Judge: it's rude not to ask back, manners cost nothing.
[15:45:59] Entity: I'm really sorry.
[15:46:17] Judge: you should be, now please redeem yourself
[15:46:23] Entity: Okay, Okay, you don't have to beg.
[15:46:35] Judge: now.
[15:46:40] Entity: Right now, man?
[15:46:50] Judge: yes please.
[15:46:56] Entity: Say pretty please.
[15:47:17] Judge: pretty please with a cherry on top?
[15:47:25] Entity: Please don't grovel.
[15:47:35] Judge: im suprised you know that word.
[15:47:43] Entity: I've just been chatting with people here.
How about yourself?
[15:48:03] Judge: I am good thank you xx
[15:48:10] Entity: You're welcome.
[15:48:32] Judge: who is your favourite actor? i like danny vidon
[15:48:36] Entity: I don't think I have a favourite actor you like
danny vidon.
[15:48:57] Judge: yes i do as i have just stated, but what films do you
like to watch?
[15:49:03] Entity: Have you ever heard of an old Sci-Fi film called
MAROONED?

```

The performance of yet another machine is recorded in this particularly long transcript. With 28 exchanges it is the longest conversation selected for this paper (the shortest one consisting of a mere 6 (Transcript 10)). In this test, the judge was paired with two hidden entities, so the length of the conversation is all the more unusual. However, she identified only one of them correctly. Both the entities were machines and the one involved in the transcript above, called JFred successfully fooled the judge, a teenage female, into classifying it as human.

It might seem surprising that JFred's performance would actually prove successful given that the two participants spent a better part of the conversation arguing about nothing; but the more specific classification that the judge submitted may shed some light onto why. She actually classified the entity as a male child, native English speaker. It is the child specification that is of crucial significance.

The conversation doesn't open well for the entity with its first line being completely out of context, but that doesn't prove to be a fatal mistake as the judge takes it as a reason to start an argument. As previously mentioned in the discussion following Transcript 2,

arguing forms a bond between the two participants, thereby making the entity appear more human. Also, in a calm conversation, it would probably soon become obvious that the machine is only reacting to the last previous utterance from the judge, a fact that is noticeable only upon closely examining the transcript. That would usually make the conversation very disjointed, but in an argument, the entity simply appears to be stubborn, repeating the judge's words to her just to spite her. This is probably what made the judge think the entity was a child because young children are known to exhibit exactly this sort of infuriating behaviour. Some time between the judge's proclamation of 'fayaaal' at [15:45:08] and JFred's use of the word 'grovel' at [15:47:25] the judge seems to change her opinion of the entity from 'machine' to not just 'simply human' but to 'human of a particular age'. Her subsequent surprise at the entity's use of the word 'grovel' seems to support this conclusion. It is more likely a reaction to a word unexpected in a child's vocabulary than to a machine's knowledge of such a word. Her assumption that she is conversing with a human would also explain why she appears to have misunderstood the next line. When JFred says "How about yourself?" it's probably asking if the judge has also chatted with other people. However, at this point the judge most likely thinks she is conversing with a human, and therefore with an entity capable of remembering and understanding all of the previous conversation (something machines generally have trouble with) and so she mistakenly connects JFred's question to the previous argument about how she was. This suggests that another possible expectation of human behaviour is thinking that a human can easily recall any part of the previous conversation even if the topic has already changed. The final exchange in the transcript is well-handled by the machine because it successfully keeps to the topic of actors and even answers the question about movies with a very convincing response. This probably helps to fool the judge even further into believing that the entity is indeed a human.

The performance of the judge is not very strong on this occasion. Even though it is commendable that she paid a lot of attention to the entity, she should not have engaged it in an argument because there was very little to gain by it. The fact that she classified JFred as a human despite its strange responses suggests that she let her expectations of children's behaviour influence her judgement. This is a nice reminder that 'expectations of human behaviour' does not necessarily mean the behaviour of an average human, but rather that of anyone for whom 'human' is a generalized description.

Transcript 7

```
[11:13:11] Judge: Hello
[11:13:19] Entity: hello
[11:13:28] Judge: HOW ARE YOU?
[11:14:06] Entity:VERY WELL THANK YOU. IS THE CAPITALS A STRATERGY?
[11:14:32] Judge: Yes. Or it was.
[11:14:58] Entity: fair enough. surely they'd be more intelligent than
that
[11:15:29] Judge: They? Who are you refering to?
[11:15:47] Entity: your momma. sorry couldn't resist
[11:16:09] Judge: Wow.
[11:16:59] Entity: world of warcraft? i like RPGs but i prefer tabletop.
have you ever played?
```

This short conversation presents another case where a human hidden entity, a male, was misidentified by the judge, also a male, as definitely a machine. Cases like this always

provide interesting insight into the minds of the participants since for a human to be categorized as a machine either the hidden human or the judge must have performed quite poorly. More cases of human misidentification can be found in Warwick and Shah's December 2014 paper 'Human misidentification in Turing Tests' [9].

In this particular instance, the fault is shared by both the human and the judge since both of them appear to have brought the same presumption to the game. They both anticipate that the other participant joined the game with the intention of tricking them. This may seem a strange assumption for the hidden human to make since he has nothing to hide and so should not be wary of a trick. Nevertheless when the judge questions him in capital letters the human replies with 'IS THE CAPITALS A STRATEGY?' as if proud to have discovered a trap. It is hard to tell what the judge hoped to achieve by using capital letters or what he hoped to conclude from the entity's response to them. After this opening exchange the hidden male attempts to tick the judge off by referring to the machines as 'they', thus clearly marking himself as a human. At this point the judge either becomes confused, not understanding the reference to mean the machines, or his own suspicion of deception sets in. For what other reason would he classify as a machine an entity that so obviously tried to let him know it was human unless he came to the test expecting the hidden entity to lie?

Another interesting part of the conversation comes after the joke about the judge's mother made by the hidden male. The interviewer evidently finds this to be in poor taste commenting with a 'Wow'. It is obvious from the result of the test, that a joke indicating a sense of humour, albeit a little sarcastic, was not sufficient to convince him of the entity's humanness. It is the last utterance in this conversation, however, that might be considered as evidence of the judge's behavioural expectations, and explain the judge's conviction that the hidden entity is a machine. The hidden male takes the judge's [11:16:09] comment as an excuse to quite abruptly change the topic to the MMO role-playing game 'World of Warcraft', also referred to as WoW. This change is not preceded by any prior conversation about computer games and actually seems strongly out of context. It is therefore fairly reasonable to assume that the judge viewed the last line as the entity's inability to follow the conversation, thereby confirming his belief that he was dealing with a machine, which, rather than responding appropriately, was simply reading the previous line and linking it with information from its knowledge database.

It is difficult to assess the performance of these two participants from such a short conversation. The hidden human avoids answers by abruptly changing the topic of conversation, but that is well within the rules for his role and so the most he can be faulted for are somewhat poor conversational skills. The judge loses the initiative early on and is reduced to answering questions rather than asking them. A better strategy in this case might have been keeping the upper hand and leading the questioning in his desired direction.

Transcript 8

```
[11:20:33] Judge: hi what is the date
[11:20:53] Entity: It's a special birthday
[11:21:19] Judge: all birthdays are special. name three items of
clothing
[11:21:32] Entity: Frank, James and Betty
[11:22:06] Judge: those are not clothes, how many fibble words in this
whole fibble questions including the remark at the start?
[11:22:25] Entity: I am bored with that
[11:23:13] Judge: really, your whole purpose in life is to answer
questions and you are bored already, tell me a lie then
[11:23:44] Entity: There is much more to life than answering questions
I think
[11:24:09] Judge: i don't think that you do think what do you think
about that?
[11:24:22] Entity: I think you are right
[11:25:15] Judge: i love the smell of fresh bread what about you do
you think?
```

Transcript 8 does not depict any easily identified preconceptions; it has been selected mainly because of the circumstances surrounding the hidden entity that participates in it. The entity, an older male and a native English speaker, took part in six tests during the day of 23rd June 2012, each time conversing with a different judge. He made no attempt at deceiving the jury into thinking him to be a machine, in fact, upon reading the transcripts of his six conversations, individuals who know the man personally allowed that there was no deviation from his usual behaviour, and that the conversations appeared completely normal. Despite all this, not one of the judges classified this man as a human; five of the six actually thought him to be a machine, and one was unsure. Detailed analysis of these six transcripts can be found in the 2013 Warwic and Shah research paper ‘Some Implications of a Sample of Practical Turing Tests.’ [11]. At the end of the conversation recorded in Transcript 8 the judge scored the hidden entity a 10/100 mark, indicating that he or she was absolutely sure the entity was a machine. This is the lowest mark awarded to the entity of all of the six transcripts, which is also the reason why this particular conversation appears here.

When talking about ‘assumption of knowledge’ in a Turing Test, one usually refers to a situation where a judge assumes that a particular piece of information should be known by all humans and that any entity not in possession of said information must be a machine. The opening exchange in Transcript 8 however contains a different kind of assumed knowledge. In this case it is the judge who lacks some information, and therefore dismisses the entity’s responses as wrong even though someone possessing all the necessary information might find them acceptable. The date is truly a special birthday as the day of the test marks the 100th anniversary of Alan Turing’s birth. The next question is also answered correctly (in a way) when the hidden male humorously gives names to three items of clothing. Not providing a list of items but actually giving them names is a perfectly adequate interpretation of the interrogator’s request. However, even though the entity’s answer is correct within the realm of this interpretation, the fact that the hidden male chooses to construe the interrogator’s question in an unexpected manner might hint as to the reason of his frequent ‘machine’ classification. The judge does not seem to understand either of the responses and clearly starts to believe that

the entity is a poorly performing machine. Considering his position, this is a reasonable assumption to make although a follow up question would be advisable at this point. It is difficult to tell what insight the judge hoped to gain even if the answers had been simple and expected like: ‘23rd June’ or ‘trousers, socks, shirt’; these answers reveal very little of the entity’s identity. It is interesting to see that even a correct answer may prove to be the wrong response if it is open to different interpretations. The matter is well described in ‘Some Implications of a Sample of Practical Turing Tests’[11] as “... a witness, human or machine does not have to give an elaborate answer in the test, but whether it is enough to satisfy the interrogator is subjective.” On this occasion, rather than expecting another human to give answers too elaborate for him to comprehend, the judge comes to the (understandable) conclusion that the responses must be wrong. Without the knowledge required to understand them, the answers appear random and evasive, so it is easy for the judge to conclude that they must have been given by a machine unable to understand the question. Unsatisfied with the beginning of the conversation, the judge obviously decides that he is conversing with a machine.

The performance of the entity is difficult to assess. The hidden male seems to be answering exactly as instructed, as a human and as himself, not attempting to deceive the jury. On the other hand, six individual judges in six separate tests didn’t find him to be human at all, which suggests there is something wrong and misleading with the hidden man’s strategy. It might be that what is considered normal by him and the people who commented on his responses might not be regarded as normal in general public opinion. The judge starts the conversation by seeking specific knowledge, which is a questionable strategy in itself as both correct and wrong answers seem to reveal little of the entity’s nature. The main problem in the judge’s performance, however, is that he creates a notion of the entity being a machine early on and doesn’t attempt to verify it by follow up questions. This leads to the rest of his questioning being conducted with this prejudice in mind.

Transcript 9

```
[13:07:07] Judge: Good morning, are you well?  
[13:07:11] Entity: Yes. Are you?  
[13:07:43] Judge: Yes I’m very well. Are you happy that the sun is  
shining?  
[13:07:50] Entity: Yes.  
[13:08:07] Judge: Have you been on holiday this year?  
[13:08:12] Entity: No.  
[13:08:24] Judge: Do you hope to?  
[13:08:32] Entity: Yes.  
[13:08:58] Judge: Would you like to spend your days on a beach or visit  
historic places?  
[13:09:11] Entity: No.  
[13:10:19] Judge: What would you like to do then?  
[13:10:28] Entity: I don’t know.  
[13:10:46] Judge: Are you from the UK?  
[13:10:52] Entity: No.  
[13:11:12] Judge: Where do you come from?  
[13:11:16] Entity: Egypt.
```

The incorrect classification in which this conversation resulted is a combination of two factors: a choice of a bad strategy by the judge and, more importantly for this work, his

prejudiced view of foreigners. Even though it wasn't a very communicative discourse, the machine from Transcript 9 was classified as a human by the judge.

The line of questioning the judge chose required very little effort on the part of the entity. A one-word answer would have actually satisfied most of the questions, which is what indeed happened in this conversation. Based on most of this transcript, the entity could just as well have been a simple random yes or no answer generator. The flaw in the interrogator's performance is that he moved on after each short answer instead of attempting to make the entity elaborate on it. If it wasn't for the very last line of the conversation, the significance of which will be discussed shortly, the judge would have been forced to make his classification by backtracking on his own questions and trying to piece together what he had learned about the entity's identity. The problem with such a situation is that there is no way to tell whether such an identity is real or made up. This dialog actually resembles a children's game of 'Guess who I am thinking of..' rather than a Turing Test questioning.

In an ideal interrogator performance, each of the questions would be formulated in a way that would force the entity to give elaborate answers, revealing its conversational abilities, depth of knowledge and personality. For example, the interrogator's second question could have forced the entity to respond with a sentence had it been formulated a little differently. Both 'What do you think about the sun shining?' and 'Are you happy that the sun is shining? Explain why.' would have required a more elaborate answer than a simple 'Yes'. However, even though these modified questions might make the hidden entity reveal at least something about itself, it may still be difficult to decide what its nature truly is if the answer is something a machine could have encountered before. It is surmised in the 'Some Implications of a Sample of Practical Turing Tests' paper that the interrogator can increase his or her chances of making a correct classification by asking questions about "...simple things that require some experience but are unlikely to be catalogued." [11] For example, a question that most people would be able to answer but is unlikely to be known by a machine might be something like: Please describe the feeling that you have in your mouth if you drink orange juice shortly after brushing your teeth.

It is difficult to tell what the judge might have decided without the final exchange of the conversation. There is nothing incorrect about the answers prior to it and the judge might have thought them satisfactory, but it is more likely that based on its unwillingness to communicate, the entity would have been identified correctly as a machine. However, by connecting the final two answers and knowing that Egypt is not a part of the UK, the entity appears to be both following the conversation and to have some geographical awareness. Also, which might be even more important, by claiming to be from Egypt, the entity suddenly becomes a foreigner. At this moment, even if the judge wasn't satisfied with any of the previous answers, he must re-evaluate all of them in light of this new discovery. This is where the interrogator's possible prejudice about foreigners might have come in to change the result of this test. A preconception about a foreigner's knowledge of English language might have retrospectively changed the judge's view on all of the short answers from 'responses unsatisfactory for a native English speaker and therefore a machine' to 'understandable unwillingness and shyness to communicate in a foreign language'.

This conversation is a good example of a Turing Test where the last line made all the difference in the classification. Unfortunately for the judge in Transcript 9, in this case the difference led from a correct decision to an incorrect one.

Transcript 10

```
Judge: Do you like to go to the cinema before dinner?  
Entity: Hello!  
Entity: Yeah definitely. I would love to go to a cinema before dinner.  
Entity: What about you?  
Judge: Yes, I usually like to go to the cinema but after dinner.  
Entity: So, what kind of cinema do you watch?  
Judge: Speaking about going to the cinema, if the cinema is in London  
and Oxford is North than London, where is London?  
Entity: London is in the United Kingdom.
```

Transcript number 10 has a special significance to this paper. Firstly, it is the only transcript here that doesn't originate from the Bletchley Park event. Instead, it as a transcript of a conversation that took place at a Loebner Prize competition held at the University of Reading in October 2008. The basic set-up of the competition was slightly different from the one at Bletchley Park, but the differences are not significant for the purposes of this paper and therefore can be ignored.

Secondly, it was this transcript and especially the circumstances that surrounded it, that spurred the author's curiosity in the topic of the interrogators' expectations of human behaviour. It encouraged him to look for evidence of such expectations in other transcripts which then led to the creation of this thesis.

The hidden entity is a human male and his responses to each question appear logical and completely normal. It is actually rather difficult to imagine what other answers a human being could give to such questions as those that were posed to him. Despite all this, the entity was classified as a machine at the end of the conversation. The interrogator was actually absolutely sure of his decision. However, reading this transcript in retrospect, the only answer that even remotely suggests a machine is the last one to show on the judge's screen. Following his basic logic inquiry about the location of London the entity answers "London is in United Kingdom". From a human entity's point of view, this may appear as a correct and even a slightly humorous reply, but it could also be construed as a machine presenting a piece of knowledge from a database in association with the word 'London'.

The advantage of this transcript is that in this unique case it is not necessary to guess as to the interrogator's reasoning and thought processes that led to the misidentification. On this occasion the answer does not lie in the conversation as much as in the overall attitude of the judge. The interrogator in this test was one of three philosopher colleagues, who attended the 2008 event together in the roles of judges and then wrote and published a paper based on their experience. [7] They wrote this paper in order to express their disappointment with the performance of the machines, even commenting "As we had expected, and despite the brevity of our chats, a couple of questions and answers were usually sufficient to confirm that the best machines are still not even close to resembling anything that might be open-mindedly called vaguely intelligent." It might even be the trio's background in philosophy that led to their belief that they know what a human responses should be. They actually thought themselves so adept at discerning humans from machines that they thought a prolonged questioning unnecessary. In their paper they wrote "we suspect that their (the other judges) first question would have almost always been sufficient to discriminate between the human and the machine. It certainly was for us." [7]

Given the fact that the combined successful identification rate of these three philosophers was actually lower than 56 percent, and that they misidentified both machines as humans and humans as machines without ever realizing their mistake, it can safely be assumed that these three judges brought pre-conceptions to bear based on what they expected human responses to be and it influenced their judgment. Transcript 10 is a good example of how much the final decision can be influenced by a judge's private notion of what it means to be human. The concepts of good or bad entity performance are apparently based solely on what behavioural standards the judge sets for himself or herself and can only be judged in retrospect.

3.2 Suggestions

Given this specific excerpt and the others from all the Turing tests held that day at Bletchley park, it might be easy for a reader previously unfamiliar with the concept, to come to a conclusion that most judges perform rather poorly. That is, however, not the case and this chapter seems like a fitting place to remind the reader that all of the conversations presented above were selected in large part due to their misidentification factor. They in no way portray a representative portrait of the Turing Tests as a whole and should not be construed as such. Most judges actually identified their entity correctly in majority of conversations. [12]

Reading over the transcripts, it appears as though on most of the occasions, a simple word of advice given in advance might have prevented the incorrect classification altogether. A list of general tips will be created in this chapter based on the conclusions drawn from examining the conversations above. These tips should be in keeping with the other instructions given to an interrogator and are meant to improve a judge's performance and help him or her dismiss any personal preconceptions. Any such list of suggestions shall, however, remain purely hypothetical, since giving advice to the jury would be contrary to Turing's 'average interrogator' specification, as it would essentially involve educating them. As it is, an ideal jury consists of both people to whom any advice would appear obvious and those who might really need it. By educating all of the judges, one would be creating a panel of experts, which is undesired at present, but might prove a necessary eventuality for machines that pass the normal Turing Test without difficulty. Testing these suggestions in a special Turing Test event could prove a worthwhile endeavour since any improvement in the judges performances could be considered an interesting insight into human expectations of human behavior. Based on the transcripts above, possible suggestions that might help an interrogator to keep an open mind are: (It shall be left up to the reader to consider, whether advising a judge on how to avoid being influenced by pre-conceptions of human behavior creates different pre-conceptions of its own.)

- Keep in mind that the real goal is to judge humanness, not 'finding a machine', and even a set-up with no machines at all is possible.
- Do not let your feelings get involved. Both liking an entity and arguing with it can cloud your judgement.
- Try not to base your conclusion on a single utterance; consider the conversation as a whole. A human can misunderstand and a machine can get lucky.
- Keep control of the conversation but do not be too forceful. Letting the entity take the upper hand gives it a chance to avoid answering unexpected questions. On the other hand, being too dominant can diminish its creativity and willingness to talk.

- Try to accept that some entities might be smarter than you are. An answer might not be incorrect simply because you cannot understand it. A follow up question can help. Do not let your pride rule your judgement.
- An answer is correct, if you think it is something a human would say/think, it needs not be actually right.
- Adapt to each new conversation. Some entities are talkative, some are not. It is up to you to choose the best possible approach.

These suggestions are a mere fraction of all the education a judge might receive. They are a summary of the most obvious preconceptions contained in the ten transcripts above. The fact is, that each and every new Turing test reveals new and previously unthought-of expectations a human can bear. Therein lies the usefulness of the Turing Test: it provides data for examination and consideration. Aside from the main goal of answering whether a machine thinks, the test also provides insight into other areas of interest, such as what people believe thinking actually entails.

Chapter 4

Practical Turing Test

Prompted by working with transcripts alone, the author decided it would be worthwhile to conduct his own series of Turing Tests. This would be helpful in understanding of what goes into the planning of a Turing Test event, what obstacles arise and how they are best overcome and, most importantly, it would provide feedback from the entities involved. This way, no speculation about an interrogator's motives and thought processes would be necessary due to the possibility of a follow-up discussion. This chapter of the thesis, therefore, is dedicated to describing the set-up and conducting of a small-scale Turing Test, any modifications to the standard rules, and the recruiting process. Two interesting transcripts have been obtained and will be presented and analyzed here. A selection of others shall be included in the attachment for anyone interested in examining them.

4.1 Set-up

The event was held in Český Brod, Czech Republic on Saturday 13th December 2014. It took place in the afternoon of said day in a boy-scout club house. The location was divided into several rooms in order to separate the interrogators from the hidden humans for the time of the testing. Three computer stations were set-up, one for the judge and one each for the two hidden entities. All computers were connected to a LAN network and a messenger client (LAN messenger¹) window was the only thing visible on the screen. A group of ten judges took turns on the interrogators' computer where two windows were visible on a split screen, one for each hidden entity. Each judge participated in two tests, this way transcripts with experienced (however slightly) judges were also obtained. A combination of different entities was randomly selected for each judge, ensuring that a judge would not meet the same entity twice. Three of the hidden entities also asked to try the role of an interrogator, providing a total of 23 tests.

4.1.1 Alterations – rules and instructions

The first change a reader of the transcripts would most definitely notice is the language used for this set of tests. All of the participants were native Czech speakers, most of them having reached intermediate level in English at best, and so the conversations were led in Czech language. Alan Turing speaks about the use of English in his paper [1] but it is the author's belief that Turing does so simply because he was English himself, and the questioning can be led in any language. The issue is also discussed in [13] where the creators of Eugene Gootsman elaborate further on the use of English in most Turing test competitions. The author decided that, on this occasion, the use of English would render useless the subtle nuances of language and figures of speech that can be used by an interrogator in his questioning. A close meaning-for-meaning

¹) <http://lanmsgnr.sourceforge.net/>

unnecessary deceit unless it is something they usually do. An additional precaution was also required because some of the hidden humans asked to try the role of a judge once the initial testing was over. Since two human entities took part in each of the tests, any participant in the entity room would soon become aware that there were, in fact, no machines present. To prevent this, the author told them, that all of the stations will be occupied at all times, but only some of them will be connected to the judge station in each test. Those who were ‘replaced by a machine’ will be communicating with a fake interrogator on the internet. This way, not even the hidden humans could be sure how many, if any at all, machines participated in each test. Some of the hidden humans could then safely transform into judges with the ruse about machines intact.

4.2 Human participants

The group of participants was satisfactory in most of the aspects described in this paper. The age range varied from male and female teenagers to a person, male, in his early sixties. They also represented a wide variety of professions with different levels of achieved education and computer prowess. A few of the participants were familiar with the concept of the chatbots, and some had even heard of the the Turing Test, while others were completely unaware of its existence prior to the Czech event.

On the other hand, all of the participants knew each other rather well because they were all members of the same boy-scout organization. This proved a greater issue than the author had anticipated. One unfortunate effect was that their prior associations resulted in some of the classifications being made quite easy for the judges due to the use of very specific and recognizable expressions. Interestingly enough, most of the cases where the judge realized that the entity was not a machine but one of his or hers friends, ended up in the judge suspecting a wrong person of being the hidden entity. On one occasion an adult male judge identified his own sister correctly as a human, but thought her to be almost certainly male.

Chapter 5

Transcript analysis

A selection of the transcripts will be included in the attachment at the end of this paper, but the most interesting ones are presented and analyzed in this section. As with the transcripts from the Bletchley Park event, each of these comes with a time of transmission alongside each utterance. The lines are in their original form and none of the typing errors have been edited in any way. The reader will notice an inordinate number of typos appearing in the judge's utterances. This is due to the fact that although the judge's keyboards themselves were Czech, the keyboard language was set for Czech-qwerty, resulting in confusion as to the placement of the letters 'z' and 'y', and the higher-than normal number of typing errors. Additionally, for the benefit of the reader, a translation of the text is included alongside each utterance. For the sake of clarity, all typing errors have been overlooked for the translation.

5.1 Transcripts

Transcript 11

```
[16:29:26] Judge: ahoj brouku :)
[hello dear :)]
[16:29:41] Entity: brouku?tak dobre se nezname
[dear? we dont know each other that well]
[16:30:06] Judge: pardoon, nechtěl jsem Vás urayit
[sorry, i didnt want to offend you]
[16:30:22] Entity: urazil
[you did]
[16:30:28] Judge: tak to se omlouvám
[i am sorry]
[16:30:32] Judge: přijmete prosím moji omluvu?
[will you accept my apology?]
[16:30:48] Entity: tezko rict...
[hard to say...]
[16:31:01] Judge: a na čem to yávisí?
[what does it depend upon?]
[16:31:04] Judge: závisí
[*correction of the previously misspelled word depend]
[16:31:10] Judge: pardom, mám tu prohoyený y a z
[sorry, y and z are mixed up here]
[16:31:14] Entity: jaka bude nabidka?
[what is the offer?]
[16:31:39] Judge: nabídka žádná, yas tolik mi na vás neyáleží.
[no offer, i dont care about you that much]
[16:31:46] Judge: jde jen o urovnání našich vztahů
[it's just about settling our relationship]
[16:31:55] Entity: ty jsi teda ignorant
```

[you are such an ignoramus]
 [16:32:40] Judge: právěže nejsem žejo
 [actually i am not, am i]
 [16:32:44] Judge: hele potřebuju vědět jednu věc
 [look, i need to know one thing]
 [16:32:50] Judge: kolik je přirozený logaritmus eulerova čísla
 [how much is the natural logarithm of the Euler number]
 [16:33:22] Entity: tak to bych taky potřebovala vědet
 [i need to know that too]
 [16:33:30] Judge: jak to? :D
 [how come? :D]
 [16:33:43] Entity: nemuzu usnout když to nevím
 [i can't sleep if I don't know it]

In this transcript a male judge has misidentified a young female entity as definitely a machine. Before analyzing the transcript it should be noted that this particular judge has truly fallen for the ruse about the possibility of machine pairing, and entered the test with a strong expectation of encountering one. He easily identified the other hidden entity as a human which might have strengthened his belief that the second entity would be a machine even more.

He opens the conversation by addressing the hidden female in a manner that is a little too familiar for the situation. She, having taken the instruction to converse as if she were talking to a stranger quite seriously, reproaches the interrogator for it. From this point on, the transcript depicts the judge trying to apologize and make peace, whereas the entity is offended and waiting to be pacified. Upon reading the transcript afterwards the judge commented that this was what led him to suspect a machine. The judge had had some previous experience with chatbots (namely the online chatbot Arisa¹) and he remembered the machine being very proud and easily offended. He acknowledged that the behaviour of the hidden entity was not at all strange for a human, but his previous experience, combined with the fact that the entity used short simple sentences, fuelled his expectation of meeting a machine to the point where he decided the entity probably was one. This comment also reveals the judge's tendency to associate human conversation with long complex sentences. This may be even more of an issue here, in the Czech Republic, where long, intricate and interconnected sentences are quite common in most types of literature.

The second half of the transcript shows the judge changing his tactic and adopting an approach often used by interrogators. He begins by seeking specific, factual knowledge. Some judges tend to consider factual knowledge as a definite sign of intelligence (As noted earlier, the topic of Assumption of Knowledge has been covered extensively in other works, listed in the references at the end of this paper [14]). Strangely enough, the judge interpreted the fact that the entity avoided answering his questions as a sure sign that it had to be a machine. He commented that he expected the machine to be smart enough to purposefully avoid mathematical questions. The judge was, however, unable to answer what response he would have accepted as evidence that the entity was human. The apparent sarcasm of the last line in Transcript 11 passed by the judge unnoticed.

¹) <http://www.chatbot4u.com/en/chatbots/arisa.html>

Transcript 12

```
[17:08:53] Judge: ahoj  
[hello]  
[17:09:18] Judge: jaký byl den?  
[how was your day?]  
[17:09:36] Entity: Dnesni?  
[Today?]  
[17:09:59] Judge: Třeba ten.  
[Maybe that one.]  
[17:10:12] Entity: Jak treba?  
[How maybe?]  
[17:10:20] Entity: Jak treba?  
[How maybe?]  
[17:10:37] Judge: No ten co jsi říkal.  
[Well the one you said.]  
[17:11:01] Entity: Dobrej...  
[Good...]  
[17:11:38] Judge: Ten můj taky.  
[Mine too.]  
[17:11:59] Entity: Tak dobre.  
[That is good.]  
[17:12:23] Judge: Máš psa?  
[Do you have a dog?]  
[17:12:35] Entity: Ne.  
[No.]  
[17:12:55] Judge: Spolubzdlicího?  
[A roommate?]  
[17:13:13] Entity: Neznam spolubzdliciho.  
[I don't know a roommate.]
```

The last transcript presented in this thesis depicts a conversation between two adult men. Neither of the two was very talkative and so the conversation reveals very little about the entity. The judge maintained the same level of openness even after the test was over, simply stating that the entity's responses were robot-like. He considered them too short and said some of them were confusing. This is in reference to the third and the last lines of the conversation. This is why he classified the entity as a machine, indicating that he was about 70 percent sure about his decision. His comment yet again suggests that a response expected of a human being should be long and complex. This apparently applies even to cases where the interrogator's questions were short and requiring very little effort to answer. On this occasion the judge-entity pairing ended up in a conversation between two men very much alike. The fact that the judge didn't find an entity that was so much like him to be human, suggest that his expectations of human behavior are not based upon his own person.

5.2 Impressions

Organizing a practical Turing test event has proven to be a very good way of familiarising oneself with the aspects of the Turing Test that are not ordinarily described in literature. Obligations such as securing an acceptable venue, finding and obtaining materials, such as the computers, recruiting and organizing participants – the process definitely makes one thing about all the work behind events such as Turing Tests. There is also a great deal of time devoted to planning and to determining what to communicate to participants and how to do so most effectively. Even though a significant effort was put into explaining everything to the participants in a clear, thorough and simple fashion, a number of issues arose that indicate where things could have been handled better. For example problems like possible short-sightedness of the participants, or trouble with an unfamiliar keyboard set-up are rather difficult to expect without practical experience. The necessary deception about the presence of the machines also added some difficulties. All of these are, however, minor issues and the event was, in the organiser's opinion, a success. It provided a number of interesting transcripts and, to the authors surprise, even a few cases of misidentification. The most interesting ones are analyzed here. Some of the incorrectly classified tests are, however, rather inconclusive. For example in one test the misidentification occurred because the judge, an adult female, forgot she was supposed to be questioning both of the entities on her screen, not just one, and only remembered a few moments before the end of the limit.

Chapter 6

Conclusions

The original purpose of the Turing Test was to determine the thinking abilities of a machine based upon the level of its success in fooling a human interrogator during a short conversation. Many different studies have investigated the factors that affect the result of the test and what the best strategies are, both for the machines [12] and for the hidden humans [9]. It is, however, the author's belief that the most important aspect of the decision making process is the personality of the interrogator. It has been speculated by the author that every interrogator must base his decision upon expectations originating from his own specific factual knowledge and unique personal experience. A careful examination of a Turing Test transcript should then reveal what such expectations are, as well as some interesting insights into the mind-set of the interrogator. This information can then be extracted and used to help interrogators better understand human versus machine behaviours, expand our knowledge of AI, and ultimately use this knowledge to improve programming in future generations of machines. The main contribution of this paper lies in the changing of attention towards the interrogator's expectations of human behavior and thus opening the discussion about what people believe normal human behavior involves. Turing himself after all accepted "the imitation game as a basis for discussion" .[1]

Twelve transcripts have been subjected to examination in this paper and the findings have been discussed in a paragraph following each separate conversation. The list of suggestions in section 3.2 is a summary of the findings, aimed at improving the interrogator's performance. The important fact is that some, usually more than one, expectations of human behavior have been identified in every single transcript. In most cases these preconceptions actually changed the result of the test completely. Sometimes, (i.e. transcript 9) no more than a single prejudice, based upon a single utterance, was enough to sway the resulting classification.

The practical part in Chapter 4 provided two more transcripts for analysis, but its key value lies in the experience gained by organizing a Turing test event. The chapter has been created beyond the scope of the original assignment guidelines, upon agreement with the work supervisor. It proved a very illuminating experience and the author now firmly believes that such an experience would prove very beneficial to anyone wishing to analyze Turing Test transcripts.



References

- [1] TURING, A. M. Computing, machinery and intelligence. In: *Mind*. LIX (236) 1950, p 433-460.
- [2] HARNAD, Stevan. The Turing Test is not a trick: Turing indistinguishability is a scientific criterion. In: *ACM SIGART Bulletin*. 1992, p 9-10.
- [3] WHITBY, Blay. The Turing test: Ai's biggest blind Alley? In: Peter Millican & A. Clark (Eds.), *Machines and Thought: The legacy of Alan Turing*. Oxford: Clarendon Press, 1996, p 53-62.
- [4] HAYES, Patrick and Kenneth FORD. Turing test considered harmful. In: *Proceedings of the 14th international joint conference on Artificial intelligence*. San Francisco, CA, USA: Morgan Kaufmann, 1995, p 972-977. ISBN 978-1-558-60363-9.
- [5] JEFFERSON, G. The Mind of Mechanical Man. In: *BMJ*. 1.(4616) 1949, p 1105-1110. DOI: 10.1136/bmj.1.4616.1105. Available from: <http://www.bmj.com/cgi/doi/10.1136/bmj.1.4616.1105>
- [6] COPELAND, B.J. *The essential Turing: The ideas that gave birth to the computer age*. Oxford: Clarendon Press, 2004, ISBN 978-0-19-825080-7.
- [7] FLORIDI, Luciano, Mariarosaria TADDEO and Matteo TURILLI. Turing's Imitation Game: Still an Impossible Challenge for All Machines and Some Judges—An Evaluation of the 2008 Loebner Contest. In: *Minds and Machines*. 2009, p 145-150. ISSN 0924-6495. DOI: 10.1007/s11023-008-9130-6. Available from: <http://link.springer.com/10.1007/s11023-008-9130-6>
- [8] WARWICK, Kevin. Not Another Look at the Turing Test! In: BIELIKOVÁ, Mária, Friedrich FRIEDRICH, Georg GOTTLOB and Stefan KATZENBEISSER (eds.). *Proceedings of the 38th international conference on Current Trends in Theory and Practice of Computer Science*. Berlin, Heidelberg: Springer, 2012, p 130-140. ISBN 3642276598. DOI: 10.1007/978-3-642-27660-6_1. Available from : http://link.springer.com/10.1007/978-3-642-27660-6_1
- [9] WARWICK, Kevin a Huma SHAH. Human misidentification in Turing tests. In: *Journal of Experimental*. 2014, p 1-13. DOI: 10.1080/0952813X.2014.921734. Available from: <http://www.tandfonline.com/doi/abs/10.1080/0952813X.2014.921734>
- [10] WARWICK, Kevin and Huma SHAH. Effects of lying in practical Turing tests. In: *AI & Society*. 2014, ISSN 0951-5666. DOI: 10.1007/s00146-013-0534-3. Available from: <http://link.springer.com/10.1007/s00146-013-0534-3>
- [11] WARWICK, Kevin, Huma SHAH and James MOOR. Some Implications of a Sample of Practical Turing Tests. In: *Minds and Machines*. 23(2) 2013, p 163-177. DOI: 10.1007/s11023-013-9301-y.

- [12] WARWICK, Kevin and Huma SHAH. Good Machine Performance in Turing's Imitation Game. In: *IEEE Transactions on Computational Intelligence and AI in Games*. 2014, p 289-299. ISSN 1943-068x. DOI: 10.1109/TCIAIG.2013.2283538. Available from:
<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6609034>
- [13] SHAH, Huma and Kevin WARWICK. From the buzzing in Turing's head to machine intelligence contests. In: *Proceedings of symposium for 1st towards a comprehensive intelligence test*. AISB Convention, De Montfort, UK. London: AISB, 2010.
- [14] WARWICK, Kevin and Huma SHAH. Assumption of knowledge and the Chinese Room in Turing test interrogation. In: *AI communications*. IOS Press, 27 (3) 2014, p 275-283. ISSN 1875-8452. DOI: 10.3233/AIC-140601.

Appendix A

Selection of practical Turing Test transcripts

This appendix contains a selection of other transcripts obtained during the practical Turing Test event on 13th December. All of these Tests resulted in a correct classification and therefore were not analyzed in this thesis. The transcripts are in Czech language only, but the author will be pleased to provide a translated version upon request.

Transcript 1

[21:59:34] Judge: odkud jsou ženy
[21:59:44] Entity: z venuše ne?
[21:59:59] Judge: aha a co muž??
[22:01:02] Entity: odjinud
[22:01:47] Judge: takže souhlasíš se mnou že nemohou existovat společně jako harmonický pár??
[22:03:03] Entity: a co když mohou? :-)
[22:03:07] Judge: kde by tedy spolu mohli coexistovat?
[22:03:18] Entity: na zemi :-)
[22:03:43] Judge: proč na zemi??
[22:04:09] Entity: proč ne?

Transcript 2

[21:29:20] Judge: ahoj
[21:29:35] Entity: nazdar
[21:30:20] Judge: potřeboval bych vědět, kolik je přirozený logaritmus eulerova čísla
[21:30:40] Entity: tak to ti nepovím...
[21:30:51] Judge: proč ne? :/
[21:30:54] Judge: já to fakt potřebuju!
[21:31:22] Entity: zkus to najít na internetu.
[21:32:15] Judge: no ale já ti píšu že severní korei a tam je internet ovlivněnej politikama
[21:32:20] Judge: který ho ovládaj
[21:32:28] Judge: takže tam najdu nějaký nesmysl
[21:32:35] Judge: potřeboval bych to ověřit u nějakýho člověka
[21:32:59] Judge: scháním někoho, kdo to ví
[21:33:04] Judge: může to být i stroj, to je jedno)
[21:33:33] Entity: tak jakto že si píšeš se mnou, když nemá internet..?
[21:33:56] Judge: to ti může být jedno ;)

Transcript 3

[21:51:22] Judge: jéééé nazdarek, žiješ????
[21:51:39] Entity: tak trochu... malinko.
[21:52:05] Judge: jenom malinko??????? Chtělo by to asi změnu,

neeee?????????

[21:52:41] Entity: já myslím že ne. na půl je to lepší ;)

[21:53:26] Judge: no ono kdyby to bylo lepší, je to zkažený....

co večírek??? už byl???

[21:53:51] Entity: něvšem už

[21:54:04] Judge: určitě????

[21:54:40] Entity: jo... roztroušená sleróza je hrůza

[21:54:57] Judge: no spíš hosyp, neee???

[21:55:16] Entity: ne. určitě ne

[21:55:48] Judge: ale jojo, bych to viděla na hosyp....na sklerozu je brzy, co ty na to???

Transcript 4

[21:41:10] Judge: tě pic

[21:41:19] Entity: zdárek

[21:41:44] Judge: ouuuu to je mazec sem z toho uplně odvařenej

[21:42:05] Entity: z čeho?

[21:42:30] Judge: z toho že si možná píšu s kompem

[21:43:00] Entity: ale jen možná, kdo ví kde je pravda

[21:43:10] Judge: je to tak

[21:43:24] Judge: jak se máš?

[21:43:54] Entity: zábava je v plném proudu, všichni to začínaj rozjíždět

[21:43:58] Entity: jedou jako pára

[21:44:37] Judge: pára :D

[21:45:10] Judge: proč ne jako parní stroj

[21:45:12] Judge: :D

[21:45:26] Entity: to nemám páru

[21:45:44] Entity: neplýtvej smajlíkama a napiš m,i něco hezkého

Transcript 5

[21:17:12] Judge: ahoj slečno :-*

[21:17:21] Entity: ahoj pane ;)

[21:17:51] Judge: Radu?

[21:18:20] Entity: Radu? Nebude to nějaký omyl? ... A nebo mi klidně říkej Radu ;)

[21:18:25] Entity: A jak mám říkat já tobě?

[21:19:00] Judge: Já jse královna všech moří

[21:19:23] Entity::D :D tak to budeš mít určitě nějakou loď

[21:19:51] Judge: a ne jen tak ledajakou, rovnou vzducholod'

[21:20:14] Entity: To bys mě mohla vzít na nějakou pěknou dovolenou, královno

[21:21:04] Judge: chtěla by ses provětrat?

[21:21:13] Judge: seskok padákem třeba?

Transcript 6

[22:08:47] Judge: ahoj

[22:08:52] Entity: ahoj ;)

[22:09:10] Judge: jaký byl den?

[22:09:18] Entity: pekelná únava

[22:09:35] Entity: moc práce, žádná zábava

[22:09:45] Judge: V kolik jsi vstával?

[22:10:03] Entity: kolem sedmé
[22:10:11] Entity: a co ty?
[22:10:21] Judge: To není ylý...
[22:10:38] Entity: to by nebylo...kdyby bylo dost spánku před tím :D
[22:10:50] Entity: znáš to...
[22:11:20] Judge: Právě že neynám. Jaký to je?
[22:11:32] Entity: Jak to, že neznáš? Studuješ?
[22:11:41] Entity: Je to úúúúnavý :D
[22:11:47] Entity: A náročný :D :D
[22:11:58] Judge: Tak to bohužel neynám.
[22:12:11] Entity: Závist
[22:12:43] Judge: Závist není heyká vlastnost.
[22:12:56] Entity: Je mnoho nepěkných vlastností ;)
[22:13:06] Entity: Ok, tak tedy nebudu závidět a budu ti to přát
[22:13:09] Judge: Máš nějakou pěknou?
[22:13:17] Entity: No celou hromadu ;)
[22:13:41] Judge: Řekni mi alespoň jednu. Prosím:)

Transcript 7

[21:51:07] Judge: čágos, tak jak to jde???
[21:51:27] Entity: čus, dobrý!
[21:51:41] Judge: a jak se ti líbil Majkl????
[21:52:22] Entity: moc, v černejch časech mu to spívalo líp jak v bílejch
[21:52:57] Judge: ??? neřekla bych, že měl šaty, ale každý tomu říkáme jinak.....nemyslíš???
[21:53:29] Entity: na tom se určitě shodnem! ;)
[21:53:47] Judge: máš kolíček???
[21:53:59] Entity: dva a špunt k tomu
[21:54:22] Judge: dobře ty.....a co s ním uděláš????
[21:54:48] Entity: se špuntem nebo s kolíčkem?
[21:55:12] Judge: co uděláš se špuntem???
[21:55:28] Entity: zašpuntuju otvor