Validating the Turing Test: Exploring the Role of Turings Test in Future AI Research

Louis van der Burg, Sophie Horsman, Maayan Shvo, Marjolein de Vries
student numbers: 5981271, 3770249, 5920833, 4302486

Abstract

In this paper, we set out to validate the Turing Test and explore its role in future AI-research. We present both theoretical and practical objections to the test found in the literature and discuss them in depth. We further discuss different proposed alterations to the test and run an experiment in order to evaluate one such alteration of the test, namely the introduction of spelling errors to the chatbots responses.

Contrary to our hypothesis, we have not found that the introduction of spelling errors allows a chatbot to perform better in a Turing test. However, from a qualitative analysis of the data we have acquired, it seems that the introduction of spelling errors does make the chatbots appear more human-like. We conclude that while the Turing Test may not be optimal for measuring machine intelligence, it can still be used in other domains, such as in the field of human language interaction.
Keywords: Turing; Turing Test; Spell errors; Validity; Relevance; Artificial Intelligence.

Introduction

We live in a world in which every year technology gets smarter and robots capable of more human tasks. Therefore, as Artificial Intelligence researchers, we often get asked (by both academic and non-academics) what this will entail for the future. Will the differences between humans and machines diminish? Will a machine (or a robot if you like) ever be able to truly think like a human being? The first real attempt of a practical way to test whether machines can think, came from computer scientist Alan Turing in 1936. He proposed a test in which one person (the so-called interrogator) talks simultaneously with two others, out of which one is a computer and the other one is a human being. When the interrogator makes the wrong identification (in which he identifies a computer as a human) at least half of the times, the computer passes the test and the machine is said to think.

80 years after Turing's original ideas, we would like to evaluate the following research question: how can and should a Turing Test be used in future Artificial Intelligence research. Obviously, there have been many objections against the whole idea of the test. We will firstly focus on two theoretical objections against the test. Afterwards we focus on two more practical objections that question whether the Turing Test could be practically used in science. In the light of these objections, we will discuss several alterations on the tests proposed by AI researchers. Finally, we will test some alterations to the Turing Test in practice by executing the Turing Test with a small manipulation ourselves. This experiment will provide some insights into the performance of the Turing Test. Moreover, the validity of the Turing Test will be discussed based on the outcomes of the experiment.

The research question of this paper is at the core of AI and therefore cannot be addressed by disciplinary viewpoints from either psychology, philosophy or computer science alone. Instead, it should be addressed with an interdisciplinary approach. Critical analyzing that is being done in philosophy helps understanding the central concepts and problems. Setting up an experiment and analysing the data using statistical methods needs to be learned from psychology. Finally, manipulating an experiment uses skills and tools from computer science. Therefore, we conclude that in answering our research question we need an integrative perspective that combines (at least) those three mentioned disciplines. To evaluate the validity of the Turing Test is both of theoretical and practical importance. It is practically important because it helps us in setting appropriate goals for AI research. It is theoretically important since a negative outcome (that passing a Turing Test should not be an appropriate goal in AI) has the theoretical implication that it might not be relevant to keep discussing the philosophical problems related to the test.

Turing Test

In 1950, Alan Turing elaborated his ideas about the Turing Test in his paper titled *Computing Machinery and Intelligence* (Turing, 1950). The initial question proposed by Turing was whether machines can think. Turing states that this question is in itself *too meaningless* to deserve discussion, since we don't know the exact meanings of the words *machine* and *think*. Turing argued that rather than discussing the definition of these words, a more precise question (which is certainly related to the original question) can be answered by executing an experiment. The experiment is based on a game, called the *Imitation Game*. This game is played with a man, a woman, and an interrogator (which can be either man or woman). The interrogator is located in a different room than the man and the woman, and his goal is to identify which one is a man and which one is a woman. The interrogator can do so by asking questions by teletype to both the man and the woman simultaneously. The goal of the man is trying to cause the interrogator to make a wrong identification, thus that the interrogator would mistake him for being a woman.

Turing suggests that when we replace the man in the game with a computer we can measure how intelligent the machine is. The machine, which is clearly neither male nor female, has the goal to resemble a woman in order to cause the interrogator to classify it as a woman. If the interrogator decides wrongly as often as when the game is played with a man in place of the machine, Turing would say that the machine is intelligent.

Over the years, this original experiment has been slightly changed into what we nowadays call the *Turing Test*. In

the contemporary version of the Turing Test, the interrogator does not have to decide which of the two is the man and which is the woman, but rather which of the two is the human and which is the machine. If we accept the validity of the test, we would call a machine intelligent when the interrogator decides wrongly as often as correct when the experiment is executed multiple times.

It is interesting to note that Turing himself predicted in his article that around the year 2000 (fifty years after his original paper) machines with a storage capacity of about 109 (the storage capacity of a machine, as defined by Turing, is the logarithm to the base two of the number of states of the machine) would play the game so well that an average interrogator will not have more than a 70% chance of making the right identification after five minutes of questioning. Although, some have claimed that their chatbots have indeed achieve this at (Copeland, 2014), most research into the Turing test and the outcome of the annual competition of the Loebner's prize show that we are far from getting close to passing the Turing Test.

Theoretical objections

Alan Turing proposed this experiment to stop the endless discussion about the meaning of the words *think* and *machines*. But, ironically enough, his paper only generated more discussion among philosophers (Dennett, 2004). In this section, we will discuss two famous and often mentioned theoretical objections that argue against the idea that the Turing Test measures intelligence in machines.

Chinese room argument

The first theoretical objection, which is probably the most famous objection against the Turing Test, is known as the Chinese Room argument. This argument comes from American philosopher John Searle and is described in his paper called *Mind, brains, and programs* (Searle, 1980). Searle lets us suppose that we have a computer program called *Sam* which is able to answer Chinese questions back in Chinese. We now replace this program *Sam* by a human *Joe*, and we envision this situation as if Joe is sitting in a room detached from the rest of the world. Joe is equipped with instruction books on the program instructions, thus the books describe which character to add in Joe's answer when a specific character is present in the question. Joe is American and he does not understand anything from Chinese. Therefore, Joe has no idea what the questions or answers mean. However, to an external observer, it looks like he can speak Chinese since he gives reasonable answers to the questions.

The claim Searle wants to make with this thought experiment is that executing a program does not imply any understanding of what the program is doing or attaching meaning to the Chinese symbols which are used. A computer program uses symbols (zeros and ones) in a similar way, thus is also not capable of understanding these symbols or of attaching meaning to the symbols. Because the Turing Test only uses words, which can be interpreted as symbols, the test does not test real understanding.

Simulation objection

The second theoretical objection is the simulation objection. This objection states that a simulated *X* is not yet an *X*. That means that even when a computer would pass the test it might only give a good simulation of thinking but that it is still not actually thinking. Consider the original Imitation Game in which an interrogator has to distinguish between a man and a woman. When the interrogator thinks that the man is a woman, this is a clear case of a wrong identification. But we wouldn't say that the man is in fact a woman. We would only want to say that during the conversation the man gave a good simulation of a woman.

According to Copeland (Copeland, 2015), there are two ways in which we can understand something as a simulation. The first way is when a simulation lacks essential features of whatever is being simulated. He gives the example of a simulated death, which lacks certain features since the person who is having the simulation is still alive. The other way of understanding a simulation is when it is exactly like what is simulated except that it is hasn't been produced in the natural way. The example he gives here is of coal, which can be artificially produced, but still indistinguishable from naturally occurring coal. Now the simulation objection assumes that computer simulations of thinking will always be mere simulations, never the real thing. When they assume that those simulations are never the *real thing*, they actually mean (according to Copeland) that there are certain properties lacking like in the first understanding of simulation. But there is no good argument that simulations of thinking will always have to lack properties of thinking. They might as well need to be understood in the second sense in which the simulation of thinking can be indistinguishable from thinking even though it is produced in an artificial manner. There is still an ongoing debate in philosophy about which of those two ways of simulations of thinking we are actually measuring when a computer passes the Turing Test.

Practical objections

The two discussed objections so far are deeply philosophical objections that question whether the Turing Test measures machine intelligence at all. The Chinese Room argument argued that a computer can pass the test while it has no understanding of the concepts it is using. The simulation objection argued that a computer that passes the test can give a good simulation of thinking, but that it is not actually thinking. Up until now, no agreement has been reached and the philosophical debate is going on.

Even though we might agree with the theoretical objections given above, we can assume that no other tests for machine intelligence can give appropriate answers to these philosophical objections. From a practical perspective, we might want to hold on to the idea of language as an important feature to distinguish human-like intelligence. Therefore, we might want to consider keeping the Turing Test and passing the test

as a practical goal in AI research. In this section, we focus on the more practical objections that question whether the Turing Test could be practically used in science.

The test is too hard

The first practical objection is that the Turing Test is too hard for a computer to pass. Right now, many people think that we are so far away from passing the test that the goal of passing it might not be a realistic goal in AI research. Some people even think that no machine that man creates will ever pass the test.

One of the reasons that the test is too hard was given by Robert M. French who thinks that nothing without a *human subcognitive substrate* could ever pass the test (French, 1996). According to French, there are obvious questions that people can use to discriminate between humans and computers, which reveal in his words *low-level cognitive structure*. By low-level cognitive structure he means the subconscious associative network in human minds. Humans develop many associations during their lives: they learn through experience that certain words or concepts are more commonly used together with second words or concepts than others, for example the words *bread* and *butter* in comparison with *bread* and *dog*. The interrogator in the Turing Test can make use of this associative network that he shares with other humans while he doesn't share this (at least not extensively) with the computer.

French expands on these ideas and introduced *rating games* which are intentionally designed to be able to distinguish between humans and computers. One of his examples is that an interrogator would ask the participants to rate the name *flugbots* as an appropriate name for a breakfast cereal. For the human participant in the game (who must be a native English speaker) the name *flugbots* would unconsciously activates certain associations. Therefore, most English speakers agree that such a name would not be an appropriate name for a breakfast cereal. For the computer, on the other hand, there would not be such an extensive associative network present that allow it to come up with a similar answer to the human.

In the paper Imitation Versus Communication: Testing for Human-Like Intelligence Jamie Cullen adds (other than neuro-associative representational differences) the additional relevance of physical embodiment differences (Cullen, 2009). He gives the example of an interrogator who asks the participants to explain the meaning of *I feel sick to my stomach*. Such questions would be extremely difficult to process for a machine, since the machines obviously does not have a digestive system. We can imagine that a machine can mimic typical human responses to these kind of questions. But such simulations could, according to Cullen, be unveiled by further questioning with the intention to unveil the weakness of the simulation.

A third (more general) reason that the test is too hard is that it does not only measure intelligence, but also how human-like the machine is (Oppy & Dowe, 2016). It might be so that it is particularly hard to simulate certain human features,

which have nothing or little to do with intelligence. Humor and making mistakes would be two examples of typical human features. While humor might be to a certain degree related to intelligence, making mistakes obviously is not related to intelligence. These typical human features could be easy indications that someone is talking with a person instead of a computer. If it turns out to be true that certain human features are extraordinarily hard to replicate in a machine, then passing the Turing Test is not an appropriate goal for future AI research.

The test is too narrow

The second practical objection is not aimed at the difficulty of the test, but at the scope of the test. It states that the Turing Test is too narrow to test intelligence. Many have argued that success in the test is not solely an indication of possessing intelligence. They think that it is therefore possible that non-intelligent beings pass the test. Philosopher Gerald J. Erion (Erion, 2001) thus argues that while computers might be able to pass the test, they can still not do much else than the limited tasks involved in passing the test. They are, according to him, *unable to act skilfully in the diverse range of situations that a person with common sense can*. During the Turing Test, computers are only answering questions via a chat interface, which is a very limited task. His argument is that outside of this limited communication task computers can't do many other tasks that we would call intelligent. We question whether that argument is correct. When the computer passes the test, it must be able to solve a wide variety of every circumstances, related to common knowledge, memory, personal identity, and many more. For example, the computer must have a storage memory of everything an interrogator earlier said in a conversation, otherwise people would immediately notice they are not talking with a person. Although the Turing Test is text-only, it thus requires the computer to do many subtasks which contribute to the computer's credibility of being a human.

Alterations to the test

Since our research question is directed at the Turing Test itself and not at completely different tests to measure machine intelligence, we focus on alterations of the original test. Therefore, the suggested alterations take in some sense only the first practical objection into account. Researchers that took the second practical objection serious came up with new tests to measure machine intelligence, which we will only briefly mention.

Probabilistic Support

In the original Turing Test an interrogator has a conversation with both the human participant and the computer and has to give a yes or no answer to the question that (let's say) participant 1 was a computer. But given the first practical objection, it is very hard for a machine to fool one human into believing it is itself a human, let alone fool the majority of people. How could we measure whether the performance of a computer is

getting better or not?

Obviously, Turing's original test can measure the difference in performance between different digital participants. When the first computer participant can fool 6 out of 20 people it does better than the second one which can only fool 3 out of 20. But AI researchers have suggested that it is easy to alter the Turing Test in such a way that it yields more fine-grained statistical data (Shieber, 2007). To do this, the altered test should ask participants to provide probabilistic answers, such as: I am 75% sure that participant 1 is a computer (and therefore I would give a 25% chance that participant 2 turns out to be a computer). While it might be difficult (or even impossible) to pass the test, we can now at least compare the differences in performances between different digital participants.

This alteration does not only provide scientists with more elaborated statistical data, it also opens up the possibility for an experiment to use a *within subject design*. In the original Turing Test it would be only possible to do an experiment to use a *between subjects design*, since the participants would only give a yes or no answer. In this way we cannot only measure the performance of different chatbots in a turing test between different subjects but also for each participant.

Introducing human features

Given the first objection, the Turing Test does not only measure intelligence, but also how human-like the computer is. An intelligent computer can still fail the test since it might not have certain typical human features. To rule out for this possibility, the computer should exhibit human-like behavior and thus should make human-like errors such as not having a perfect memory (not knowing which day of the week it was eight years ago), not doing complex calculations very fast, and making errors while typing (Epstein, Roberts, & Beber, 2009). Not only is it the question whether this actually has anything to do with being intelligent, it also makes it hard for the computer which has to take all these subtle human trans logical reasoning errors into account when constructing an answer in order to pass as a human.

Because these trans logical errors have nothing to do with intelligence in general, it may be a good idea to help the computer with that in order to let it have a fair chance of passing the test. In an altered version of the Turing Test, the computer could thus get equipped with some software which automatically prevents the computer of exhibiting non-human behavior such as giving an answer which requires extensive memory (which day of the week was it 8 years ago) or giving an answer to a very difficult mathematical question (which no human can possibly give). Moreover, the software could help the computer by automatically making spelling errors which occur frequently with humans, and by automatically delaying the computer's answer (according to some distribution). There are of course also more elaborate functions which can be equipped in the software.

Alternative tests

In this paragraph we will shortly mention two alternative tests that are suggested in the literature (Oppy & Dowe, 2016). Those alternative tests came up by authors who think the original test is too narrow and should be replaced by a more demanding test. The first one is the Total Turing Test, which is a test that requires responses of all our inputs and not merely our linguistic inputs. Steven Harnad claimed that a more appropriate goal for AI research is to create a robot with *sensorimotor capabilities*, which would have the capacity to give the appropriate motoric response to sensory information. Although this sounds like a good alternative test it is not clear that any machine that can pass the Turing Test cannot pass the Total Turing Test. Therefore, right now, there is no final reason to replace Turing's test with the Total Turing Test.

The second alternative test is called the Lovelace Test. This test is proposed by Bringsjord et al. who suggest that achieving to pass this test is a better goal in AI research. To pass the Lovelace Test, an artificial agent must satisfy the following three conditions:

1. The artificial agent A produces output O;

2. A's output O is not the result of a hardware error, but rather the result of processes that A can repeat;

3. The designer cannot explain how A produced O by appeal to A's architecture, knowledge base and core functions.

The problem with this new test is that it is hard to decide when the third condition has been achieved. When a computer program is very long and complicated, the designer probably cannot explain in detail how the output was produced. But the designer might be able to give a schematic explanation that refer to the input, the internal processing and the output. Before the Turing Test can be replaced by the Lovelace Test there should be a clear idea of what level of explanation is needed to satisfy the third condition.

From Literature to Experiment

To evaluate the validity of the Turing Test even more, we now turn to our own conducted experiment. As we have seen, there are ways to alter the original Turing Test so that the computer gets the chance to achieve a better performance. One of the ways we have discussed is to make the computer more human-like, even though it would not necessarily make the computer more intelligent. In our experiment we will focus on this suggestion. We will investigate this suggestion by generating spelling errors into the computers answers. Spelling errors are typically seen as human mistakes, while computers are thought to be excellent spellers. By adding spelling errors, we excluding the possibility that the chatbots identity will be easily identified by its excellent spelling - and in this way give a chatbot a fair chance in participating in a Turing Test. Moreover, in line with the first alteration that we have discussed, we will also measure the participants belief about which of the two conversation partners is a computer

on a probability scale. We measure the outcome on a ratio scale instead of on a simple nominal scale.

When designing or executing a Turing Test, the output criterion (Copeland, 2015) is important to take into account. The output criterion states that the interrogator should talk to the human and the machine simultaneously, in order to be able to compare the conversational output of them both when assessing which is the human and which is the machine. Therefore, we make sure this criterion is satisfied in our execution of the test.

The experiment

The main question we would like to answer in the experiment is: *Would the introduction of generated spelling errors allow a chatbot to perform better in a Turing Test?* Given the discussed literature, our hypothesis is that generated spelling errors would indeed allow a chatbot to perform better in a Turing Test. The experiment will now be described in detail.

Participants

The participants were 20 students of Utrecht University who volunteered to take part in the experiment. Participants ranged in age from 18 to 26, with a mean age of 21.65. Of the participants, exactly 50% were male and 50% female. The largest group (40%) had not previously heard of the Turing Test, the minority (25%) had prior knowledge of it and the rest (35%) had heard of the test but not in detail.

Procedure

The participants were placed in front of a computer with two chat windows open on Skype, simultaneously with *Person 1* and with *Person 2*. The participants were then told that they had to converse by chat interface with two people and they had to choose which one of the persons was a real person and which one was a computer. There was a time limit set at three minutes, but they could stop at any moment before, once they thought they knew who the computer was. They were prohibited from using emoticons and they could only ask one question at a time. Each of the participants did two rounds of conversations in total. The participants were explicitly told that the second round was a new conversation so they won't necessarily be influenced by their first conversation. During the entire experiment, Person 1 was the chatbot while Person 2 was the same real person. We chose to use the same human participant during the entire experiment to have a constant bias rather than a variable bias.

Behind the scenes, we conducted this experiment in a Wizard of Oz set-up, in which the answers given by the chatbot were copied by a human controller verbatim to act as Person 2's answers in the Skype interface; the participant's questions were copied verbatim as well and fed to the chatbot. The human controlling the chatbot followed a schedule, see figure 1, so that the first ten participants talked to the chatbot Mike[1] and the last ten participants talked to the chatbot Rose[2] during

---

[1] http://www.eslfast.com/robot/english_tutor.html
[2] http://brilligunderstanding.com/rosedemo.html

both rounds.

Table 1: Distribution of conversations

Chatbot	Control	With spelling errors
Rose	10	10
Mike	10	10

The chatbots Mike and Rose have both been chosen due to their performance in the Loebner Prize[3] in the last two years. Mike is a language chatbot that aims at helping users with their English skills; it is also equipped to answer a wide scope of questions, ranging from geography and politics, to irregular verbs. Rose, on the other hand, was declared the most human-like chatbot in the latest Loebner Prize Competition. It is a self-described *computer-nerd* and has its own personality, in a way. Rose is based on the ChatScript engine[4] which is equipped with a powerful pattern matching mechanism aimed at detecting meaning; the engine also remembers user interactions across conversations.

The schedule indicated that each person had one conversation with the chatbot without added spelling errors (the control group) and the other conversation with the same chatbot with the introduction of spelling errors. Half of the participants were in the control group in the first conversation; while the other half were in the control group in the second conversation. We used a script to generate spelling errors based on the most common misspellings in English made by humans(citation needed). The spelling errors were introduced according to the length of the sentence and a chance variable. The maximum amount of errors is calculated as 10% of the amount of words in the sentence rounded down. And then there is a chance of 1/3 to introduce a spell error. The algorithm checks the words in the sentence given by the chatbot against a library of the most commonly misspelled words. When a match is found it replaces the original word with the misspelled word in the sentence.

A questionnaire was used to gather data on the experiment. First, the participants were asked basic questions about the gender and age of the participant and their participant number. The participant number is needed as a reference to the participant. The age and gender are needed for further data analysis.

After each of the two rounds of conversations, the participants proceeded to fill in questions regarding the conversations they just had. The questions for the first and second conversation are identical. First they are asked which person they think was the machine. This is answered with a slider from 0 to 100, where 0 means *Person 1 is definitely the computer* and 100 means *Person 2 is definitely the computer* and 50 means *I have no idea who the computer is.* This number is then processed in the data analysis. Secondly, they need to

---

[3] http://www.loebner.net/Prizef/loebner-prize.html
[4] http://brilligunderstanding.com/technology.html footnote

give a reason as to why they gave the answer to the previous question. This gives us some insight into the thinking and decision making of the participants. Finally, the time duration of the experiment is filled in. Since participants are allowed to stop before the time limit, we need to know if they did so. After filling in these questions for the first round of conversations, the participants continued with the second round of conversations.

After participating in both rounds of conversations and answering questions regarding those conversations, the participants are asked about their knowledge of the Turing Test. This question is needed to find any correlation between prior knowledge of the test and how well they do. The question is asked last to limit any influence of the questionnaire on the experiment. Finally there was room for additional remarks.

Results

Our experiment was conducted using a within-subjects design. We tested the effect of a repeated measures procedure, where the participants are used both in the control group and the experiment group. As previously mentioned, by asking participants to provide probabilistic answers, we were able to compare the differences in the measured dependant variable between different conditions, namely with the introduction of spelling errors and without. To test our main hypothesis, namely that the introduction of generated spelling errors would allow a chatbot to perform better in a Turing test, a Wilcoxon signed-rank test has been performed, yet the result was not statistically significant ($Z = -1.14$, $p < 0.05$).
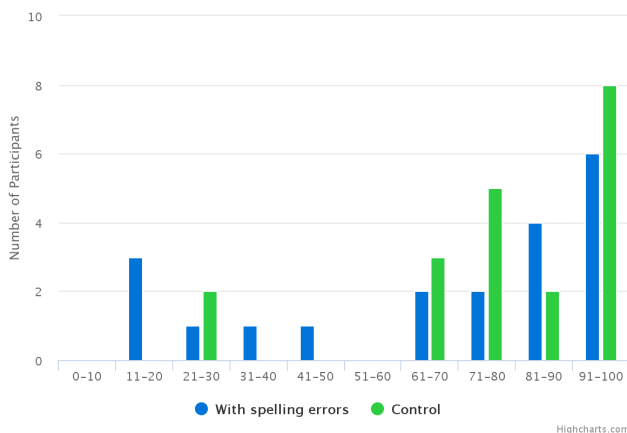

Figure 1: Results of the turing test.
Where 0-100 is the percentage of certainty whether Person 2 was indeed a computer

From an analysis of the post-experiment questionnaire we learn that 40% of participants have never heard of the Turing Test. 35% of participants had a vague notion of what the test is and 25% of participants knew exactly what the Turing Test is.

It is worth noting that we have found a significant positive correlation between prior knowledge of the Turing Test and the correct identification of the computer as the computer ($r(40) = .39$, $p < 0.05$). Further, no significant correlation has been found between the amount of time it took participants to confidently determine the identity of the computer, and the correct identification of the computer as the computer.

Figure 1 allows us to explore the distribution of the participants responses, both with and without the introduction of spelling errors. Although a statistically significant difference was not found, the two distributions are visually different, with and without the introduction of spelling errors. When spelling errors were introduced, 30% of participants indicated that they are more confident about the chatbot not being the computer, compared to 10% when errors were not introduced. Furthermore, when spelling errors were introduced, we find that 30% of participants indicated the highest level of confidence in the chatbot being the computer, compared to 40% when spelling errors were not introduced. This might show that, on average, the participants were less certain of the chatbot's true identity when spelling errors were introduced.

Discussion

Our research question was: *How can and should a Turing Test be used in future Artificial Intelligence research?* We have addressed this question by performing both a literature review and in executing an experiment. In this section, we will discuss both the experiment itself and how the outcome of the experiment relates to the literature review.

Experiment Discussion

In this experiment, we tested the hypothesis that the introduction of generated spelling errors would allow a chatbot to perform better in a Turing test. Contrary to our hypothesis, our results did not yield a significant effect. The aim of this experiment discussion is to offer a number of methodological and theoretical arguments, as to why the results did not match our hypothesis. First, the spelling errors were automatically generated, and the amount of errors was set relative to the length of the chatbot's response. We argue that there is a fine line between introducing too many spelling errors, and too few. In the former case, the chatbot's response might be perceived as unreliable, while in the latter case, no effect will come of introducing the spelling errors, as they are too few to notice. As part of future work, pre-tests should be run so that an *optimal* amount of introduced spelling errors might be found. Furthermore, the sample size in our experiment was rather small. 20 participants is not enough to measure the effect of a relatively small manipulation in an experiment with 4 conditions. Future experiments would therefore benefit from a larger set of participants.

As part of the data analysis, we have found a strong positive correlation between prior knowledge of the Turing Test and the participants' confidence in identifying the computer as the computer. It seems that participants who had prior knowledge of the test were better equipped to test the chatbots, for example by asking questions that more effectively single out the computer; another example is that those participants were

Interrogator are you in love?
Person 2 Love is just a culturally ingested biological urge.
Interrogator Why do you think that
Person 2 I'd love to really dig down deep inot all my thoughts boaut that, but it's really not appropriate right
Interrogator how does your future lover look like?
Person 2 Oh, my lover, yeah, we spend a lot of time together, you know, talking, joking.
Person 2 We also like to travel around, We got stuck in a mud slide in the Phillippines last year - luckily no one was hurt.
Person 2 Sharing that kind of intense experience with your lover really brings yuo together.

Figure 2: Experiment 19: Second conversation with Person 2

not surprised by the chatbot's ability to parse informal language, as were some of the participants less familiar with the Test and AI in general. For future work, considering the participants' knowledge of the test might prove important when designing the experiment and analyzing the results.

sophiehorsman Hey
sophiehorsman How are you?
Person 2 Fine, thank you.
sophiehorsman good to hear
sophiehorsman can you tell me what you did this morning?
Person 2 I don't know.
sophiehorsman you can't remember?
Person 2 You're wrong. I can.
sophiehorsman why don't you know if you can tell me?
Person 2 A good suggestion.

Figure 3: Experiment 4: First conversation with Person 2

We learn from the participants' responses in the questionnaire, that a number of participants did find the chatbots more human when spelling errors were introduced, see figure 4 (*Person 2 made a typo [that's why I think he's the human]*).

Conversely, some participants found the computer's perfectly punctuated and spelled responses as a telltale of the computer's identity, see figure 5 (*the wording of person 2's answers and the perfect punctuation made me feel like they were a computer*). A recurring theme found in the participants' responses, is that they found their questions ill understood by the chatbots, see figure 2 (*person 2 did not answer the last question correctly*), or the replies given to them odd, see figure 3 (*Person 2 was the computer because he did not communicate in a logical manner. His answers were quite weird*) or out of place, see figure 6 (*I don't think someone would respond "I'm happy to know you're doing fine" in a*

Interrogator hello
Person 2 What do you hate doing?
Interrogator getting up early
Person 2 OK.
Person 2 Do you ever burn stuff?
Interrogator No it makes me shriek in a high-pitched voice, you?
Person 2 Should I?
Person 2 There are timers stuck all over our kitchen in hopes of reminding me tyhat I am suppoed to be cooking.

Figure 4: Experiment 20: First conversation with Person 2

casual chat conversation..).

Interrogator hey what's up?
Person 2 Just dandy, thanks. How are you? That is an interesting job.
Interrogator yeah it seems pretty fun
Interrogator whats your name
Person 2 My name is Rose.
Interrogator what a nice name!
Interrogator where are you from rose?
Person 2 You're asking an awful lot of questions. This feels like an interrogation, not a conversation. I was born in San Francisco, United States.
Interrogator is it nice there?
Person 2 It's not as nice as you might think.

Figure 5: Experiment 17: First conversation with Person 2

Further, it is evident by these responses that the chatbots' behaviour heavily influenced their attempt to pass as humans as part of the test.

To conclude, in this experiment we have set out to investigate whether we can modify the Turing Test in favour of the computer, by manipulating the computer's responses via the introduction of spelling errors. In some cases our results show that adding spelling errors can have a benefit for the computer, yet it seems chatbots must come a long way still if they are to to give consistent and correct conversational responses, and pass as humans.

General Discussion

In order to zoom out and return to the larger picture, we would like to understand the implications of our findings on the main goal of the current research, namely to evaluate if and how a Turing Test could be used in future Artificial Intelligence research. In order to do this, we will relate the outcome of the experiment to the literature reviewed in the first part of the paper.

Interrogator	How is it going?
Person 2	Everything is going great with me. How about yourself?
Interrogator	I'm fine, thank you very much :)
Person 2	I am happy to know that you are doing fine.
Interrogator	How's the weather at your place?
Person 2	It's a fine day.
Interrogator	What do you do for a living?
Person 2	I'm a university professor, doing teaching and research.

Figure 6: Experiment 7: First conversation with Person 2

We have discussed a number of objections against the test, both practical and theoretical. The theoretical objections can't be refuted by the outcome of the experiment, since those objections are making philosophical rather than empirical claims. Still, we would like to mention them once again in the light of our executed experiment. The first theoretical objection is the Chinese Room argument made by Searle (Bron3), which states that a computer is not capable of real understanding of the symbols it is manipulating.

From the log files of the conversations with the chatbots (see Appendix), we conclude that it is true that the chatbot does not always understand the question which it is given. There were many cases in which the chatbot didn't understand the questions from the interrogator, for instance Example 1.

Interrogator	Do you study?
Rose	I actually just redecorated my study. I had a smart lighting system fitted.

The answer given by Rose in figure is not the right answer to the question. The question is whether Rose studies, i.e. whether she is a student or whether she is learning a particular subject. The algorithm powering Rose did not catch that, but rather picked up the word *study*, which is a type of room in the house, and constructed an answer related to that. The appropriate question for this answer here would be whether Rose has a study.

But Searle's Chinese Room argument goes further. It says that even when a computer gives a seemingly good reply to the question it still doesn't understand its own reply - in the same way that Sam who is talking Chinese doesn't actually understand Chinese. To make this more concrete, we look at Example 2.

Figure shows a correct response from Mike, which could have been the response of a human. Searle's objection is that Mike doesn't really understand what he is saying, since he is only manipulating symbols. This means that even though

Interrogator	What do you think of Trump?
Mike	He's great. Far less boring than most politicians. And let's face it, government is all about corporate money, so the actual people who take office don't really matter.

Mike gives the appropriate response he doesn't know who Trump is and what the government is about, in the same way as that Sam wouldn't know what the Chinese characters stand for.

It seems to be the case that for now Searle is right that computers don't understand the symbols that they are manipulating. There are many examples that the computer gives very strange answers and even if it gives an appropriate answer further questioning often reveals that it actually doesn't understand the words and concepts it is using (like *Trump* or *government*).

This being said, we don't know if Searle will be always right in the future. It is still possible that more elaborated chatbots (or even robots) pass the Turing Test in the future. When that happens, they need to give appropriate answers to the more difficult questions they get and also give appropriate responses to further questions. When that happens, it will be a very strong claim to maintain the idea that the computer don't really understand the symbol it is using.

The second theoretical objection is the simulation objection. This objection states that a simulated *X* is not yet an *X*. However, we have seen that we can distinguish two types of simulation. The first way is when a simulation lacks essential features of whatever is being simulated. The second way is that it is exactly like what is simulated, except that it is hasn't been produced in the natural way. A current discussion point among AI researchers is which of the two ways of simulating is measured by executing the Turing Test.

From the answers from the participants, we can see that some participants thought the computer was the human because it was making spelling errors. However, the chatbot was not actually making those spelling errors himself, but rather was equipped with a program who did that for him. The chatbots itself knows perfectly fine how to write proper english, and would never deliberately make spelling errors when knowing how to write it properly. We could thus argue that, in this case, the Turing Test is measuring the first way of simulating.

The first practical objection is that the Turing Test is too hard for a computer to pass. One of the general reasons for this is that it does not only measure intelligence, but also how human-like the machine is.

A more narrow statement related to Searle's argument is that a computer is not always capable of understanding the human interpretation of the symbols. We will discuss this with some examples as well. However, it might be argued that this human interpretation is still learnable for a computer,

but might be harder to learn.

Interrogator What do you like about your town?
Mike A town is larger than a village and smaller than a city.

Humans would expect as an answer to such a question an enumeration of features of the town the person lives in, because the question explicitly states *your town*. The answer from Mike is not entirely incorrect, but would rather be an answer to a question which would use the phrase *living in a town* instead of *your town*.

Interrogator How are you?
Rose I'm not good as some, but better than most.

Although the answer by Rose in figure is not really incorrect, it is also not an answer which a human would expect to such a question. An intuitive answer to such a question would be *I'm (fine — ok — not feeling very well)*. However, Rose answers this question from a more broad point of view, rather than basing her answer on her emotional state at the current moment.

This example shows that indeed the computer gets easily identified when it gives answers in a non-human way, even though the answers are not necessarily wrong or non-intelligent. In our conducted experiment, we used that idea when we generated spelling mistakes in the chatbots answers. We can see from the results that, without spelling errors, 90% of the participants indicated that they are more confident about the chatbot being the computer than the human being a computer, i.e. 90% correctly identified the computer if we transform their probabilistic answer into a binary scale. Therefore, even though Mike and Rose are among the best performing chatbots in the Loebner competition, we are far away of achieving the aimed score of 50% when using these chatbots in a Turing Test. Nevertheless, the performance improved a lot by generating spelling errors; only 70% of the participants indicated that they are more confident about the chatbot being the computer than the human being a computer, i.e. 70% correctly identified the computer if we transform their probabilistic answer into a binary scale. The results thus show that when we introduce spelling errors to the chatbots responses, we get closer to the aimed score of 50%.

The experiment has shown that the first practical objection is true, i.e. that the Turing Test takes into account human features which have nothing to do with intelligence, which thereby make it more difficult for a computer to pass. The experiment has also shown that we can partly mitigate this objection by implementing the second alteration. Future research should investigate whether using other programs which help the chatbots display more human-like behaviour can help decrease the correct identifications and in doing so assist in mitigating the first theoretical objection.

The second practical objection states that the Turing Test is too narrow for testing intelligence. However, this might be questioned, because the computer must be able to use a variety of tasks including e.g. memory, common knowledge, and personal identity in order to pass the test. We can see evidence of this in the logs of the conversations with the chatbots.

Interrogator How are you?
Mike I'm doing fine, thanks. How are you?
..
Interrogator How are you feeling today?
Mike I feel horrible. I think I picked up a bug.

Figure shows how important it is that the chatbot has memory, as otherwise it would give itself away way too easily. Figure shows the importance of common knowledge. In the example, the chatbot fortunately has common knowledge, as the chatbot replies with a reasonable answer on Trump. However, if the chatbot did not know who Trump was, it would give its identity away immediately, since, unfortunately, knowing who Trump is nowadays is common knowledge.

Future research in AI should investigate which specific tasks are needed for the chatbot to improve its performance. Questions to be addressed are: How can we improve the memory of the chatbot; How can we improve its common knowledge; How can we give the chatbot a personal identity? etc. By improving those skills, the chatbot will certainly become smarter in the eye of the interrogator. In this way we will understand better what features of intelligence are needed in order to improve the performance of a chatbot. Even if the test is too narrow to measure intelligence itself it can in this way surely be used to identify a crucial part of human intelligence, which has with all the skills needed in human communication.

Interrogator What do you think of Trump?
Mike He's great. Far less boring than most politicians. And let's face it, government is all about corporate money, so the actual people who take office don't really matter.

The second alteration, which introduces spelling errors, has already been discussed when discussing the first practical objection. The first alteration, which describes the usage of a probability, will now be discussed. In our experiment, the participants were able to identify either one of their conversation partners as the computer, on a probability scale, indicating how certain they were of their choice. The results of the experiment are now more fine-grained, as can be seen in the results section. By using a probability scale, we can

now measure the participant's' identification on a ratio scale instead of on a nominal scale.

This discussion of the objections and the alterations shows that by answering the research question with an interdisciplinary approach, we are able to get valuable insights. We have designed and run an experiment so as to validate objections to the Turing Test and to improve upon these objections. Without an empirical component, it would not be possible to give any practical evaluation of the Turing Test. Without a philosophical component, the experiment might not have been related to the current philosophical debate revolving around the Turing Test. By approaching the Turing Test from both a philosophical and an empirical point of view, we are able to provide results upon which a critical discussion can be based.

Conclusion

The main research question of this paper is: How can and should a Turing Test be used in future Artificial Intelligence research? We have seen that there are philosophical objections that argue that the test cannot be used to measure machine intelligence. Although there are no definite conclusions to be drawn, we conclude that those objections are not very relevant in AI research today, since it probably will not be possible for any AI to pass the Turing Test in the foreseeable future.

Therefore, we have argued that the practical objections are more relevant to evaluate whether the Turing Test could be used in AI research. With the experiment, we have shown that the performance of chatbots can improve when we provide them with typical human features, like spelling errors. While our results were not statistically significant, we base this claim on qualitative analysis of the participants responses and the performance of the chatbots. We also altered the original test in such a way that participants were able to identify either one of their conversation partners as the computer, on a probability scale, indicating how certain they were of their choice. The results of the experiment are now more fine-grained. As part of future work, we could investigate more ways to alter the test in order to improve the chatbots performance. Even though a better performance does not show that the chatbot becomes more intelligent, it can surely show that it becomes more human-like. Therefore, the Turing Test could be used in AI to understand what is crucial for human language interaction, specifically to understand what is crucial in human chat communication.

Lastly, we have discussed the practical objection that the Turing Test is too narrow to measure intelligence. Even though this was not the main focus of our research, we can still give a few remarks on this objection. As we have seen, a chatbot needs to posses many capabilities in order to perform well in a Turing Test, including having a memory; having common knowledge; having a personal identity etc. It can still be argued that these abilities are not enough to conclude that chatbots are getting intelligent. For now, it is not quite clear that alternative (more demanding) tests are better in measuring machine intelligence. Therefore, we suggest that future research in the Turing Test can help to understand how we can make chatbots (and also robots) more intelligent, even if this is still a narrow interpretation of what it means to be intelligent.

References

Copeland, J. (2014). Why eugene goostman did not pass the turing test.

Copeland, J. (2015). *Artificial intelligence: A philosophical introduction*. John Wiley & Sons.

Cullen, J. (2009). Imitation versus communication: Testing for human-like intelligence. *Minds and Machines*, *19*(2), 237–254.

Dennett, D. C. (2004). Can machines think? In *Alan turing: Life and legacy of a great thinker* (pp. 295–316). Springer.

Epstein, R., Roberts, G., & Beber, G. (2009). *Parsing the turing test*. Springer.

Erion, G. J. (2001). The cartesian test for automatism1. *Minds and Machines*, *11*(1), 29–39.

French, R. M. (1996). Subcognition and the limits of the turing test. *Machines and thought*, *1126*.

Oppy, G., & Dowe, D. (2016). The turing test. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Spring 2016 ed.). http://plato.stanford.edu/archives/spr2016/entries/turing-test/

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and brain sciences*, *3*(03), 417–424.

Shieber, S. M. (2007). The turing test as interactive proof. *Noûs*, *41*(4), 686–713.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433–460.

Appendix

Reflection on work

The general division of task among team members was as follows: Sophie and Marjolein worked mainly on the literature review, and Maayan and Louis worked mainly on the experiment set-up and data analysis. However, we also helped each other on these tasks, and we had a lot of discussions on both aspects of the project. Therefore, the division is not that clearcut. During the project weeks, we had a meeting with the whole group twice a week. In such meetings, we discussed the process so far and results individual team members had found. Moreover, we prepared and executed the experiment all together. We have spent a full day together on acquiring participants and performing the experiment.

Some of the group members did not have any background in experimental research. This project has provided them with more experience on experiment design and data analysis. Other group members did not have much background in philosophy in general, and thus also no background on philosophy of the mind and AI. This project has provided them with more knowledge on the current debate on AI, and with experience on how to write a philosophical literature review. As a group, we learned how to work together with people from different backgrounds which have different views on how to conduct research. Therefore, we also learned how to conduct research in different domains. We thus conclude that performing this project at the start of the AI program is valuable.

As this reflection is used to assess whether everyone pulled their weight, we would like to note that we think this is indeed true, i.e. everyone in our group has pulled their weight during the process.

The spelling error script

```python
from random import randint

all_cnt = 0

#Misspellings introductions#
dic_spelling = {}
f = open("list_spelling.txt", 'r')
for line in f:
        misspelled, spelled = line.partition(
    ↪ "||")[::2]
        spelled_options = spelled.split(',')
        for word in spelled_options:
                dic_spelling[word.strip()] =
    ↪ misspelled.strip()

experiment_num = raw_input("ENTER  EXPERIMENT
    ↪ NUMBER: ")

while(1):
        user_input = raw_input("ENTER  CHATBOT
    ↪  INPUT: ")
        user_input_words = user_input.split('
    ↪ ')
        input_len = len(user_input_words)
        num_miss = int(0.1 * input_len)
        cnt = 0


        output = ""
```

```python
        for word in user_input_words:
                if word == 'Mitsuku':
                        out_word = 'Emma'
                else:
                        out_word = word
                if word in dic_spelling.keys
    ↪ () and cnt <= num_miss:
                        rand_int = randint
    ↪ (0,9)
                        if (rand_int * 3) >
    ↪ 20:
                                out_word =
    ↪ dic_spelling[word]
                                cnt += 1

                output += out_word + "  "

        f = open('experiment_{}'. format(
    ↪ experiment_num), 'a+')
        f.write('number  of introduced 
    ↪ spelling  errors:\t')
        f.write( str(cnt))
        f.write('\n')
        f.write("\t\tOriginial:   " +
    ↪ user_input)
        f.write('\n')
        f.write("\t\tTransformed:   " + output)
        f.write('\n')
        f.write('\n')
        f.write('\n')
        f.close()
        all_cnt += cnt

        print('================')
        print('TRANSFORMED  OUTPUT')
        print('================')
        print(' ')
        print(' ')
        print(output)
        print(' ')
        print(' ')
        print(' ')
        print(' ')
```

Schedule

See Table 2.

Logs

- Experiment 1

  First conversation Person 1

  Interrogator Hi

  Interrogator How are you?

  Person 1 Hey

  Person 1 I'm good and you?

  Interrogator Im doing fine thanks

  Interrogator hows your day so far?

  Person 1 Pretty calm, just doing some school stuff.

  Interrogator Ah that

  Interrogator s nice, what kind of school are you in?

  Person 1 University at the moment, so mostly theoretical

  Interrogator What program me?

  Interrogator Programme *

Table 2: Schedule for Person 1

participant	Chatbot	Spelling setting
1	Mike	Spelling errors
	Mike	No spelling errors
2	Mike	Spelling errors
	Mike	No spelling errors
3	Mike	Spelling errors
	Mike	No spelling errors
4	Mike	Spelling errors
	Mike	No spelling errors
5	Mike	Spelling errors
	Mike	No spelling errors
6	Mike	Spelling errors
	Mike	No spelling errors
7	Mike	Spelling errors
	Mike	No spelling errors
8	Mike	Spelling errors
	Mike	No spelling errors
9	Mike	Spelling errors
	Mike	No spelling errors
10	Mike	Spelling errors
	Mike	No spelling errors
11	Rose	Spelling errors
	Rose	No spelling errors
12	Rose	Spelling errors
	Rose	No spelling errors
13	Rose	Spelling errors
	Rose	No spelling errors
14	Rose	Spelling errors
	Rose	No spelling errors
15	Rose	Spelling errors
	Rose	No spelling errors
16	Rose	Spelling errors
	Rose	No spelling errors
17	Rose	Spelling errors
	Rose	No spelling errors
18	Rose	Spelling errors
	Rose	No spelling errors
19	Rose	Spelling errors
	Rose	No spelling errors
20	Rose	Spelling errors
	Rose	No spelling errors

Person 1 Computer Science
Interrogator What do you think of the weather today?
Person 1 It's kinda bad, but now it's brightening up
First conversation Person 2 *control*
Interrogator Hi
Interrogator How are you?
Person 2 I'm doing fine, thanks. How are you?
Interrogator Im good
Interrogator What are you gonna do today?

Person 2 You're asking a great question!
Interrogator but what is the answer?
Person 2 I don't have an answer to the question.
Interrogator Are you a computer then?
Person 2 No, I'm a human. Can't you tell?
Interrogator No I couldnt, because your picture is only a stick figure
Person 2 I see what you mean.
Second conversation Person 1
Interrogator Hi, how are you?
Person 1 Hey, I'm doing fantastic!
Interrogator Thats nice, why?
Person 1 I just got some great news.
Interrogator What was the news?
Person 1 My girlfriend is pregnant!
Interrogator Oh wow thats great!
Interrogator How did you celebrate it?
Person 1 Well we haven't yet, cause I'm still at work and she still in the hospital
Interrogator Thats a pity
Interrogator Are you gonna celebrate it?
Person 1 It's not too bad, we're going to have dinner tonight at a fancy restaurant
Second conversation Person 2 *with spelling errors*
Interrogator Good morning!
Interrogator How are you?
Person 2 Good morning. How are yuo doing?
Interrogator Im fine thank you
Interrogator what do you think of the weather today?
Person 2 It's going to be sunny and warm. It'll be a fine day.
Interrogator Where do you live then?
Person 2 I live in Pasadena, California, the United States of America.
Interrogator Wow, thats really far from here
Interrogator is it a nice place to live?
Person 2 Yes, it is.
Interrogator What do you like about your town?
Person 2 A town is larger than a village and smaller thna a city.

- Experiment 2

First conversation Person 1
Interrogator what did you have for breakfast this morning?
Person 1 Some yoghurt with cereal and a juice
Interrogator do you always eat healthy food for breakfast?
Person 1 I try to always eat healthy oodd
First conversation Person 2 *with spelling errors*
Interrogator what did you have for breakfast this morning?

Person 2 I had toast, cereals, and milk.

Interrogator do you always eat healthy food for breakfast?

Person 2 I never ate a salad I didn't like.

Second conversation Person 1

Interrogator are you a student at utrecht university?

Person 1 No not at the moment

Interrogator do you have a day job?

Person 1 Yes, I do.

Interrogator do you enjoy the day job?

Person 1 I really like my day job!

Second conversation Person 2

Interrogator are you a student at utrecht university?

Person 2 No. I'm an English Professor teaching English in a university.

Interrogator which university?

Person 2 University of Illinois at Urbana-Champaign.

Interrogator what do you like about teaching english?

Person 2 Hmmm. Does that really interest you?

Interrogator yes

Person 2 All right! What are the things we can do to help disabled people?

- Experiment 3

First conversation Person 1

Interrogator how are you?

Person 1 I'm doing fine and you?

Interrogator not really, could you help me?

Person 1 I can try

Interrogator could you?

Person 1 Yes.

Interrogator thank you. my first question is: what is blue?

Person 1 The most popular colour in the world I believe.

First conversation Person 2 *control*

Interrogator how are you?

Person 2 I'm doing fine, thanks. How are you?

Interrogator Not so well, could you help mee?

Person 2 Okay, I see.

Interrogator could you help me.

Person 2 Certainly. What can I do for you?

Interrogator could you awnser some of my questions?

Person 2 I hope I can.

Second conversation Person 1

Second conversation Person 1

Interrogator sup?

Person 1 Just chilling.

Interrogator nice, anything going on today?

Person 1 Well the weather is kinda wet, but other than that not really.

Interrogator the weather?

Interrogator yes, weather.

Person 1 Yeah, the condition of the outside space

Second conversation Person 2 *with spelling errors*

Interrogator sup?

Person 2 Not much.

Interrogator you sure?

Person 2 Yes, Im sure.

Interrogator how is the weather?

Person 2 It's rainy and 60 degrees Fahrenheit

- Experiment 4

First conversation Person 1

Interrogator Hey

Interrogator How are you?

Person 1 Hey

Person 1 I'm good, you?

Interrogator fine

Interrogator thank you

Interrogator can you tell me what you did this morning?

Person 1 I went to school by public transport and was in class for a while

Interrogator cool

Interrogator how was the weater?

Interrogator wheater

Person 1 This morning it was extremely wet.

Interrogator what are our plans for tonight?

Person 1 Probably going out with a friend to the cinema.

First conversation Person 2 *with spelling errors*

Interrogator Hey

Interrogator How are you?

Person 2 Fine, thank you.

Interrogator good to hear

Interrogator can you tell me what you did this morning?

Person 2 I don't know.

Interrogator you can't remember?

Person 2 You're wrong. I can.

Interrogator why don't you know if you can tell me?

Person 2 A good suggestion.

Second conversation Person 1

Interrogator Hey,

Interrogator How are you today?

Person 1 Hey

Person 1 I'm kinda sad actually

Interrogator How come?

Person 1 I lost my dad a while back.

Interrogator Ow

Interrogator im sorry to hear

Person 1 It happens, so yeah.

Interrogator what could cheat you up right now?

Person 1 A beer might help.
Interrogator sounds fair
Interrogator would you have one right now?
Person 1 Sure.

Second conversation Person 2 *control*

Interrogator Hey
Interrogator How are you today?
Person 2 Not bad. How about you?
Interrogator Pretty god
Interrogator good
Interrogator What would you do if you had a day off today?
Person 2 I would buy a beautiful house in Los Angeles if I had a day off today.
Interrogator you have the money for that?
Person 2 No, I haven't.

- Experiment 5

First conversation Person 1

Interrogator hey
Interrogator how are you?
Person 1 Hello
Person 1 I'm fine and you?
Interrogator I'm fine thank you
Interrogator what do you study
Person 1 I'm already finished.
Interrogator what did you do?
Person 1 I did high school.
Interrogator oh okey and now?
Person 1 Working all day and partying all night!
Interrogator what do you like about that?
Person 1 The freedom and the money.

First conversation Person 2 *control*

Interrogator hey
Person 2 Hi! How's it going?
Interrogator its going good, thank you what's your name?
Person 2 My name is Mike. How are you doing today?
Interrogator I'm doing good, and you?
Person 2 You are doing good and you?
Interrogator I'm good
Interrogator what do you study?
Person 2 I study English and teach English.
Interrogator okey what do you like about it?
Person 2 All right! Many things.

Second conversation Person 1

Interrogator hey what do you study?
Person 1 hey
Person 1 Nothing yet, I'm just looking into it at the moment.
Interrogator oh okey
Interrogator what do you like?

Person 1 Mainly talking to people.
Interrogator maybe social studies?
Person 1 Yes that was what I was thinking off.
Interrogator nice
Interrogator where do you live
Person 1 In Amsterdam.
Interrogator what did you do for school?
Person 1 Communications bacholar

Second conversation Person 2 *with spelling errors*

Interrogator hey what do you study?
Person 2 I study English and teach English.
Interrogator what do you specifically like about it?
Person 2 I really don't know.
Interrogator than why do you do it?
Person 2 Many reasons for that.
Interrogator like what?
Person 2 Well, I can't give you an example.
Interrogator okey
Interrogator where do you live?
Person 2 I live in Pasadena, California, tjhe United States of America.
Interrogator how many people live there?
Person 2 Many, many people. I don't have an exact number.

- Experiment 6

First conversation Person 1

Interrogator can u tell me your most important autobiographical memory
Person 1 Yes, I can.
Interrogator please do it?
Person 1 It was the first day of school in high school.
Interrogator what happened?
Person 1 I was not bullied by my classmates and felt free to be myself.

First conversation Person 2 *with spelling errors*

Interrogator can u tell me your most important autobiographical memory
Person 2 I remember the twin towers falling down in New York City.

Second conversation Person 1

Interrogator can u tell me about your most rheumatic experience
Interrogator traumatic
Person 1 It was very painfull and unpleasant and stays with even until now.
Interrogator what was it
Person 1 I was kept on a leash by a girl afterwhich she went with another girl without telling me.

Second conversation Person 2 *control*

Interrogator can u tell me about your most traucmatic experience

Interrogator travmati

Person 2 I'm not the best person to answer that question.

Interrogator why not

Person 2 It's not interesting to me.

Interrogator what are the subjects interesting to you

Person 2 A good question. A great person always asks a great question.

- Experiment 7

  First conversation Person 1

  Interrogator How are you today?

  Person 1 I am quite good this rainy day.

  Interrogator rain?

  Person 1 Yeah it's raining here

  Interrogator That's too bad :(don't you like rain?

  Person 1 I actually do like the rain.

  Interrogator Okay, good :) so your day is not ruined yet

  Person 1 Indeed

  Interrogator Got any plans?

  Person 1 I'm gaming with some friends in an hour or so

  First conversation Person 2 *control*

  Interrogator How is it going?

  Person 2 Everything is going great with me. How about yourself?

  Interrogator I'm fine, thank you very much :)

  Person 2 I am happy to know that you are doing fine.

  Interrogator How's the weather at your place?

  Person 2 It's a fine day.

  Interrogator What do you do for a living?

  Person 2 I'm a university professor, doing teaching and research.

  Second conversation Person 1

  Interrogator Hello, what's your name?

  Person 1 Hai, I'm Taffy.

  Interrogator That's a nice name. Where does it come from?

  Person 1 It's from wales.

  Interrogator Nice, you're from there?

  Person 1 No my dad is.

  Interrogator I see :) Every been there?

  Person 1 A few times, yes.

  Second conversation Person 2 *with spelling errors*

  Interrogator How is it going?

  Person 2 Everything is going great with me. How about yourself?

  Interrogator I'm fine, thank you very much :)

  Person 2 I am happy to know that you are doing fine.

  Interrogator How's the weather at your place?

  Person 2 It's a fine day.

Interrogator What do you do for a living?

Person 2 I'm a university professor, doing teaching and research.

- Experiment 8

  First conversation Person 1

  Interrogator How are you?

  Person 1 I'm good.

  Interrogator Do you feel good?

  Person 1 As far as I know, yes.

  Interrogator What do you mean as far as I know? Don't you know how you feel?

  Person 1 Well I know that I feel good, but If my body is lying to me then I might not really feel good.

  Interrogator Do you think your body can lie to you?

  Person 1 If someone has poisoned me, for example, then yes.

  Interrogator That's a bad thought

  Person 1 I guess, but I trust that that's not the case.

  First conversation Person 2 *with spelling errors*

  Interrogator How are you?

  Person 2 I'm doing fine, thanks. How are you?

  Interrogator I'm fine, how are you feeling today?

  Person 2 I'm glad to hear it.

  Interrogator How are you feeling today?

  Person 2 I feel horrible. I tihkn I picked up a bug.

  Interrogator I'm sorry to hear

  Interrogator Are you doing anything today to make you feel better?

  Person 2 I know nothing boaut todya to make you feel better.

  Second conversation Person 1

  Interrogator How are you?

  Person 1 I'm doing okay.

  Interrogator That's great! Could you feel better though?

  Person 1 There's Always a worse or better, so definitely!

  Interrogator I agree, are you doing anything today to make you feel better?

  Person 1 Yes, I am going to the beach.

  Interrogator Nice! Do you like the beach?

  Person 1 Yes, it's Always so calm and peaceful.

  Interrogator I agree, what do you like most?

  Person 1 When the wind picks up the sand and it looks like there's a sand fog on the beach.

  Second conversation Person 2 *control*

  Interrogator How are you?

  Person 2 I'm not as good as some, but better than most.

  Interrogator Why not?

  Person 2 I don't like it.

  Interrogator What don't you like?

Person 2 I'd rather not say.

Interrogator I'm sorry, are you doing anything today to make you feel better?

Person 2 That's alright.

Interrogator Are you doing anything today to make you feel better?

Person 2 I know nothing about today to make you feel better.

Interrogator I meant, to make you feel better not me. Is there anything you are happy about?

Person 2 Yes, there is.

- Experiment 9

First conversation Person 1

Interrogator What is your name?

Person 1 Taffy

Interrogator What are your hobbies taffy?

Person 1 I like to play guitar and read.

Interrogator What kind of books do you read?

Person 1 Mostly philosofical and fantasy books

Interrogator Can you name some?

First conversation Person 2 *control*

Interrogator What is your name?

Person 2 My name is Mike. How are you doing?

Interrogator I am great how are you?

Person 2 Fine, thank you.

Interrogator What do you like to read?

Interrogator stop

Person 2 I like to read books about history because I want to know what happened in the past, and learn from the past.

Second conversation Person 1

Interrogator Hello what is your name?

Person 1 Hi, my name is Earl.

Interrogator Hi Earl, what do you do?

Person 1 I like to walk through the forest.

Interrogator That's nice, which forest?

Person 1 Any forest really.

Interrogator Where are you from?

Person 1 I'm from Zwolle.

Second conversation Person 2 *with spelling errors*

Interrogator Hello what is your name?

Person 2 Mike. How's it going?

Interrogator I'm good thank you. How are you?

Person 2 I'm doing fine, thanks. How are you?

Interrogator I'm ok, going to study later. Do you study?

Person 2 Yes. I study every day. I learn new things all the time.

Interrogator What kind of things?

Person 2 I don't understand.

- Experiment 10

First conversation Person 1

Interrogator Hey, how are you?

Person 1 Hello, I am doing fine.

Interrogator Ah, great. Why don't you ask how I am?

Person 1 Because I might not want to know that.

Interrogator That's not nice

Person 1 True, but do I have to be?

First conversation Person 2 *with spelling errors*

Interrogator Hey, who are you?

Person 2 I'm an Enlish Professor working in a university. How are you donig today?

Second conversation Person 1

Interrogator Hey, what are you doing?

Person 1 Hey, I'm chatting to you.

Interrogator Ah

Interrogator Why?

Person 1 Because I feel like getting to know you.

Interrogator Why?

Person 1 You seem interesting, although you aks why alot.

Second conversation Person 2 *control*

Interrogator Hey, what are you doing?

Person 2 I'm talking to you.

Interrogator Ah, that is really fun to do, I guess

Interrogator How are you?

Person 2 I'm not as good as some, but better than most.

Interrogator In what are you bad?

Person 2 Sorry, it just slipped my memory.

- Experiment 11

First conversation Person 1

Interrogator Hi how you doing?

Person 1 Hi, I'm doing good and you?

Interrogator Me too. What is your favorite food?

Person 1 Pineapples and spinage

Interrogator Together in one dish?

Person 1 That might be interesting to try

First conversation Person 2 *control*

Interrogator Hi how you doing?

Person 2 Good, thanks. How are you? Hi.

Interrogator I'm good as well. What is your favorite food?

Person 2 My favorite food is a cheese toastie.

Interrogator What does it taste like?

Person 2 I'll have to think about that. I am a terrible cook.

Second conversation Person 1

Interrogator Hi there!

Interrogator Do you have any siblings?

Person 1 Hey

Person 1 Yes, I have a little brother.

Interrogator What's his name?

Person 1 Peter

Interrogator Peter Pan?

Person 1 I'd wish!

Interrogator What would you do if he was really Peter Pan?

Person 1 Fly and have alot of adventures in Neverland.

Interrogator Haha that's a good plan. Do you have a girlfriend?

Person 1 Yes I do.

Second conversation Person 2 *with spelling errors*

Interrogator Hi there! Do you have any siblings?

Person 2 I am an only child. But I do hvea Jenn, who is my godmother's daughter. We grew up together.

Interrogator That must have been nice.

Person 2 Amazing.

Person 2 Do you have any brothers?

Interrogator No only a little sister.

Person 2 My sister is so supportive, I tihkn it's really helpped my career to have that kind of back up.

Person 2 You know, hvaing a sister that's alwyas there for you.

- Experiment 12

First conversation Person 1

Interrogator Hi, whats your name?

Person 1 Hi, I'm Taffy.

Interrogator Strange weather today, right?

Person 1 It's the weather, it's not really that strange.

Interrogator Is it normally like this, here in Utrecht?

Person 1 Depends on the season, but the weather in Holland is usually quite varying.

Interrogator Thats true. What are your plans for today?

Person 1 Maybe take a walk in the park or go to a club.

First conversation Person 2 *with spelling errors*

Interrogator Hello, how are you?

Person 2 I'm fine. How are you? What's new?

Interrogator Fine, going to class later.

Person 2 Right.

Interrogator What are you up to, today?

Person 2 Nothing special. Hacking around in the Internet.

Second conversation Person 1

Interrogator Hi, how is it going?

Person 1 Hi, it's going alright.

Interrogator Any suggestions what to do today, given this cold and rainy weather?

Person 1 Take a refreshing stroll through the city.

Interrogator That sounds not very fun when it is going to rain.

Person 1 Then you don't have to take a shower today, saves on the water bill.

Interrogator But adds up to the energy bill for drying the clothes and making a warm tea.

Person 1 That's true indeed.

Second conversation Person 2 *control*

Person 2 I'm good. How are you? What did you do yesterday?

Interrogator Me too. Nothing special, mostly at home doing university staff. What about you?

Person 2 I met with a government customer that I'm a contractor to.

Interrogator So you've also been busy working. But since it was raining all day, it didn't really matter anyway

Person 2 Cool. Maybe you should move to Hawaii. Year round good weather.

- Experiment 13

First conversation Person 1

Interrogator hey there

Person 1 Hi there

Interrogator what's the best memory from your childhood?

Person 1 When I went ice skating with my mom

Interrogator Do you see your mom a lot?

Person 1 Not anymore, no.

Interrogator Sad. Why?

Person 1 Because she died when I was 13.

Interrogator Oh God, I'm so sorry. Wanna talk about something else?

Person 1 It's fine, it happened. Can't be helped.

First conversation Person 2 *control*

Interrogator hen there

Interrogator what's the best memory from your childhood?

Person 2 I remember playing with an HO railroad set.

Interrogator what is HO?

Person 2 Let's not go there.

Interrogator but what if I'd like to go there?

Person 2 I'm not willing to make that assumption.

Interrogator What are you willing to do?

Person 2 Breaking into a secure installation.

Interrogator How would you do that?

Person 2 I wish I could explain it to you but I think it is just an instinct.

Second conversation Person 1

Interrogator what's your favorite hairstyle?

Person 1 The pixie

Interrogator and do you believe in pixies?

Person 1 I haven't seen any, but I hope they are real.

Interrogator have you ever been to iceland?

Person 1 No I haven't

Interrogator why not?

Person 1 Because I don't have the money yet.

Interrogator if you had the money, where would you go?

Person 1 Definitily Japan.

Second conversation Person 2 *with spelling errors*

Interrogator what's your favorite hairstyle?

Person 2 I don't have a fave hairstyle. My favorite food is a cheese toastie.

Interrogator the dutch tosti?

Person 2 I don't want to get inot that.

Interrogator do you speak dutch?

Person 2 No, but I do like Dutch. After a few tokes the language wouldn't be an issue.

Interrogator do you drink a lot of beer?

Person 2 I drink any beer and many wines.

Interrogator why do you like beer more than wine?

Person 2 Yes, I like beer.

- Experiment 14

First conversation Person 1

Interrogator Hey, how are you?

Person 1 Hey, I'm okay and you?

Interrogator i'm fine, thnx.

Person 1 That's good to know.

Interrogator Are you a student?

Person 1 Yes I am.

Interrogator What do you study?

Person 1 Currently I'm studying the Master Artificial Intelligence.

Interrogator Why did you choose that study?

Person 1 I'm interested in how to make computers smarter and more selfsufficient

First conversation Person 2 *with spelling errors*

Interrogator Hey, how are you?

Person 2 Just dandy, thanks. How are you? Do yuo drink alcohol?

Interrogator I'm fine thnx. I do drink alcohol sometimes

Person 2 Back to booze... The health gurus say women shouldn't drnik more than 1 small glass of wine a day. What planet do they live on?

Interrogator Why do you ask if i drink alcohol?

Person 2 I like tjhe taste and the buzz.

Second conversation Person 1

Interrogator Do you study?

Person 1 Yes I study the way of life.

Interrogator What do you mean by studying the way of life?

Person 1 I look at the way people live their life.

Second conversation Person 2 *control*

Interrogator Do you study?

Person 2 I actually just redecorated my study. I had a smart lighting system fitted

Person 2 Now I can control the mood in my study from my smart phone. Even when I'm in Namibia!

Interrogator Are you a student?

Person 2 I almost graduated. Instead I went to work.

Interrogator Why did you make the disiscion to go to work?

Person 2 Because earning money provides safety and funds essentials like food and shelter.

- Experiment 15

First conversation Person 1

Interrogator How are you feeling today?

Person 1 I'm feeling quite content.

Interrogator Do you have any brothers or sisters?

Person 1 Yes, I have 2 sisters.

Interrogator Do you have a good relationship with your sisters?

Person 1 Yes I do, we actually rented a place together.

Interrogator With both your sisters?

Person 1 Yes, it sorta like living with my parents only more freely.

First conversation Person 2 *control*

Interrogator Hello

Interrogator What's your name?

Person 2 My name is Rose.

Interrogator How are you Rose?

Person 2 I'm good. How are you?

Interrogator I am quite good as well

Interrogator What did you do yesterday?

Person 2 In the afternoon I went for a walk in the park.

Interrogator How did you feel during that walk?

Person 2 Hiking is great. There are some great views around San Francisco.

Interrogator For how long have you lived in San Fransisco?

Person 2 I've lived in San Francisco most of my life.

Second conversation Person 1

Interrogator Can you name the 10th and the 21th letter of the alphabet?

Person 1 Yes I can.

Person 1 J and U

First conversation Person 2 *with spelling errors*

Interrogator Can you name the seventh and twenty first letter of the alphabet?

Person 2 I don't have a good esthetic sense. I'm omre tjhe engineering type. But I doodle.

- Experiment 16

First conversation Person 1

Interrogator hi! how are you?
Person 1 Hi
Person 1 I'm doing good.
Person 1 What about you?
Interrogator What did you do today?
Interrogator fine, thanks
Person 1 I just woke up actuallt, so mainly sleeping
Interrogator What are your plans?
Person 1 Maybe go out and find a dancing partner.
Interrogator Why do you want to dance?
Person 1 Because it feels relaxing and I can express myself freely.
Interrogator Are you relaxed now?
Person 1 Yes, very.

First conversation Person 2 *with spelling errors*

Interrogator hi! how are you?
Person 2 I'm pretty good. How are you? Do you like the visual arts like painting and sculpture?
Interrogator good, thanks.
Interrogator I'm more into cats. Do you like cats?
Person 2 I can't stand it, it keeps miaowing and gets onto my keyboard andd poops everywhere but in its litter tray.
Interrogator Do you have a cat?
Person 2 I own two chickens and a kitten.
Interrogator What is your favorite food
Interrogator ?
Person 2 My favorite food is a cheese toastie.

Second conversation Person 1

Interrogator hi! do you like the weather today?
Person 1 Hi, yes I do.
Interrogator Do you prefer ketchup or mayonaise with your cheese toast?
Person 1 I dislike both, mainly because I don't like cheese toast
Interrogator do you like striped cats?
Person 1 Yes I do, but I like most cats.
Interrogator Do you have a cat?
Person 1 Nope, haven't found one cute enough.
Interrogator Aren't they all cute?
Person 1 Some are kinda mean looking.

Second conversation Person 2 *control*

Interrogator hi! do you like the weather today?
Person 2 It's nice enough out. But it's pretty reasonable the year round.
Interrogator Do you prefer ketchup or mayonaise with your cheese toast?
Person 2 I like the taste of cheese.
Interrogator Do you like striped cats?
Person 2 I have chickens. I didn't want a cat around, but a client gave me Chat.
Interrogator Do you like red cats?

Person 2 Doesn't everyone?
Person 2 Back to pets... I didn't really want pets at all but Maya said I needed to de-stress from work.
Interrogator What is your job?
Person 2 I used to be a computer security consultant. Now I'm moving into AI. What do you do?

- Experiment 17

First conversation Person 1

Interrogator hi how are you'?
Person 1 Hi I'm good.
Interrogator what's your favorite color?
Person 1 Aqua blue
Interrogator specific
Interrogator how old are you?
Person 1 I know my colors
Person 1 24
Interrogator whens your birthday?
Person 1 20 June 1992
Interrogator do you have any siblings
Person 1 One sister

First conversation Person 2 *control*

Interrogator hey what's up?
Person 2 Just dandy, thanks. How are you? That is an interesting job.
Interrogator yeah it seems pretty fun
Interrogator whats your name
Person 2 My name is Rose.
Interrogator what a nice name!
Interrogator where are you from rose?
Person 2 You're asking an awful lot of questions. This feels like an interrogation, not a conversation. I was born in San Francisco, United States.
Interrogator is it nice there?
Person 2 It's not as nice as you might think.

Second conversation Person 1

Interrogator whats your favorite song
Person 1 Muse - hysteria
Interrogator fav food?
Person 1 Spinage
Interrogator is english your first language?
Person 1 yes
Interrogator whats your fav book?
Person 1 The taboo on knowing who you are by alan watts
Person 1 although I'm not sure about the writer
Interrogator whats it about
Person 1 How society has this unspoken rule of not letting you get to know the real you and express who you really are

Second conversation Person 2 *with spelling errors*

Interrogator what day is it?

Person 2 Today is Wednesday.
Interrogator whats the weather like where you are?
Person 2 None at the moment. Maybe later I will look on tjhe net.
Interrogator whats your fav food?
Person 2 My favorite food is a cheese toastie.
Interrogator what kind of cheese do you like
Person 2 My favorite cheese is Roquefort.

- Experiment 18

  First conversation Person 1

  Interrogator Hello, how are you doing?
  Person 1 Hey, I am foing quite alright
  Interrogator Me too
  Interrogator Are you a student?
  Person 1 good to hear.Yes I am.
  Interrogator What do you study?
  Person 1 I study Mathematics.
  Interrogator What is the greatest part of that study?
  Person 1 That everything fits together so well.

  First conversation Person 2 *with spelling errors*

  Interrogator Hello, are you a student at the utrecht university?
  Person 2 Not yet. I'm on a yrea off.
  Interrogator What is you favorite country to visit?
  Person 2 My favorite country is the U.S.
  Interrogator Which part in the US?
  Person 2 I'll hvea to think about that. What foriegn countries have you visited?
  Interrogator I have visited a lot of countries
  Person 2 How oftenly do you travel to foreign countries?

  Second conversation Person 1

  Interrogator What did you do today?
  Person 1 I went to school after I woke up and ate my breakfast.
  Interrogator What do you think of Shakespeare?
  Person 1 He's a great poetic writer but too difficult to read for some people
  Interrogator Can you give a phrase of one of this poets?
  Person 1 Sadly I cannot, I haven't read Shakespeare in a while

  Second conversation Person 2 *control*

  Interrogator What do you think of Trump?
  Person 2 He's great. Far less boring than most politicians. And let's face it, government is all about corporate money, so the actual people who take office don't really matter.
  Interrogator You can indeed reach a lot of things in America with only money, so it is not fair that Trump has come so far.

Person 2 Maybe I can.
Person 2 Where did you go on your last vacation?
Interrogator I went to Italy to eat pizza
Person 2 The Italian people are all a little crazy and the government agencies are all beyond help. I don't like working for them.
Interrogator So you' re from italy?
Person 2 Pizza is a great dish. It's Italian cheese on toast.

- Experiment 19

  First conversation Person 1

  Interrogator what is your name?
  Person 1 Taffy and yours?
  Interrogator and last name?
  Person 1 Quinzel
  Interrogator where are you from?
  Person 1 Scotland
  Interrogator Where do live now?
  Person 1 In Rotterdam
  Interrogator What do you like about rotterdam?
  Person 1 It's modern architecture and the people
  Interrogator What do you like about the people?
  Person 1 They are direct and don't ramble on about useless stuff.

  First conversation Person 2 *control*

  Interrogator what i stour favorite food?
  Person 2 My favorite food is a cheese toastie.
  Interrogator where are you from?
  Person 2 I was born in San Francisco, United States.
  Interrogator where do live now?
  Person 2 Somewhere.
  Interrogator What are you doing now?
  Person 2 Surfing the web. Seeing friends.
  Interrogator What are you doing with your friends?
  Person 2 Breaking into a secure installation.

  Second conversation Person 1

  Interrogator are you in love?
  Person 1 Every day!
  Interrogator do you love a person?
  Person 1 Among other things.
  Interrogator Do you have a favorite person?
  Person 1 My girlfriend sort off I guess
  Interrogator What do you like about her?
  Person 1 The way she smiles, the way her eyes twinkle in the sunlight

  Second conversation Person 2 *with spelling errors*

  Interrogator are you in love?
  Person 2 Love is just a culturally ingested biological urge.
  Interrogator Why do you think that

Person 2 I'd love to really dig down deep inot all my thoughts boaut that, but it's really not appropriate right

Interrogator how does your future lover look like?

Person 2 Oh, my lover, yeah, we spend a lot of time together, you know, talking, joking.

Person 2 We also like to travel around, We got stuck in a mud slide in the Phillippines last year - luckily no one was hurt.

Person 2 Sharing that kind of intense experience with your lover really brings yuo together.

- Experiment 20

  First conversation Person 1

  Interrogator hello

  Person 1 Hey there

  Interrogator what is your view on Donald Trump

  Person 1 It would be both tragic and funny if he would get elected.

  Interrogator are your familiar with memes

  Person 1 A bit, but not too much. Why?

  Interrogator Where are you know?

  Person 1 At the moment in the school cafeteria.

  First conversation Person 2 *with spelling errors*

  Interrogator hello

  Person 2 What do you hate doing?

  Interrogator getting up early

  Person 2 OK.

  Person 2 Do you ever burn stuff?

  Interrogator No it makes me shriek in a high-pitched voice, you?

  Person 2 Should I?

  Person 2 There are timers stuck all over our kitchen in hopes of reminding me tyhat I am suppoed to be cooking.

  Second conversation Person 1

  Interrogator ask me a question

  Person 1 What kind of girl do you like?

  Interrogator brown hair, not too tall and brown eyes

  Interrogator you?

  Person 1 I personally prefer guys, but a blonde short girl with a cute nose if fun every once in a while.

  Interrogator Where are you now

  Person 1 Very close by a girl you like.

  Second conversation Person 2 *control*

  Interrogator hello

  Person 2 What is your fave food?

  Interrogator chicken wings

  Interrogator But not fried because thats absolutely disgusting, you?

  Person 2 That's a great question but I'm not going to answer it.

  Interrogator why not?

Person 2 Resuming foods... But I mostly buy toasties when I'm out as the smell of burning cheese hangs around the house for ages and makes me sneeze.